

Prioritizing Australia

Australia's researchers have been set national goals by their government. The underlying agenda is to ensure that the country's institutions spend scarce funds more wisely and in a way that reflects joined-up thinking.

In a landmass nearly the size of the United States and occupied by fewer than 20 million inhabitants, Australia's research community suffers not so much from a tyranny of distance as one of scale. Australian science takes pride in wringing high-quality research from scant resources. Despite increases in government funding over the past two years, industrial support for research remains scarce, and Australia's universities are cash-strapped. Now it risks slipping further behind as the price of research rises and the United States and Britain bolster research funding to levels that dwarf Australian investment.

With its tight resource constraints, Australia must target areas of strength or greatest need. Last week the federal government announced priorities for research in science and technology (see page 597), and adopted broad themes couched with politically palatable rhetoric. The theme 'Safeguarding Australia' will resonate with a nation burying its dead from the terrorist attack in Bali and worrying about bioterrorism and susceptibility to agricultural epidemics, such as foot-and-mouth disease. Likewise, an 'Environmentally Sustainable Australia' will be balm to a country ravaged by drought, salinity and bushfires.

With the two other themes embracing health and technology, it seems that relatively few areas of scientific pursuit will miss out, especially after some expedient re-labelling of research objectives. Notable exceptions are big physics and astronomy, although enterprising efforts in instrumentation by Australian astronomers could, and should, be considered a new technology frontier.

This priority setting is a politically strategic exercise to encourage the various parts of the research sector to work more closely together using existing funds. The government's light-touch approach to nudging its research objectives may bring a sigh of relief from researchers, but the devil may be in the implementation. No apportionment of existing funds to the priorities has been specified.

Demonstrable success in priority areas will no doubt be perceived as a draw card for future funding. In this climate, research agencies must be careful not to pirate a broad basic-research base in response to the prioritization demands of their paymaster.

A measurable return on investment will be expected, and Australian researchers are not alone in trying to devise sound performance indicators for fundamental research. Traditional measures of output, such as publication record, need to be carefully handled: dependence on them as a criterion for funding is stimulating an overproduction of Australian publications (see *Nature* **419**, 877; 2002).

Furthermore, there have been expectations that industry would blossom around research investment. But venture capital and pharmaceutical investment remain weak, despite government-supported seed funds and tax incentives. The question remains whether there is a strong entrepreneurial attitude in Australia towards commercializing discoveries, and whether overseas investors can be tempted away from more familiar territories. This remains a key challenge for government, business and entrepreneurial researchers.

The priority setting is intended not only to focus the minds of researchers on problems relevant to the populace, but to encourage nationally coordinated networks of research to attain the critical mass needed to produce top-quality science. This is supported by the government's call last month for a national audit of the research sector, due for completion next year. In theory, this could help to meld the fragmented research community, and mobilize funds and efforts across research 'silos' to address common objectives. But these collaborations must be productive, rather than mere marriages of convenience to attract more funding. To that end, Australia must address the challenge of how best to foster collaboration without diminishing the competitive and independent spirit that drives its best, world-class research. ■

Through the protein labyrinth

Proteomics is delivering results that pose new problems in data management and compatibility.

Spots, bands and peaks. Readers could be forgiven for being hard pressed to make a connection but, to the up-and-coming biologist, resolving spots and bands on polyacrylamide gels and reading peaks from a mass spectrophotometer lie at the heart of high-throughput biology. Technology and robotics lead the protein-biology revolution and, as quantity gains credence, biologists must take control of the avalanche of information being presented.

Proteomics is defined as the identification and quantitative expression analysis of every protein encoded by a genome, characterizing protein-protein interactions and assigning these protein complexes with function. Biologists have undertaken large-scale screens to identify every protein expressed on a single chromosome and determine the complement of proteins that defines a complete biological process. But in the rush, the golden rules of biology that allow for easy comparison, exchange and verification of data are being overlooked.

Hordes of data have been analysed, but it is difficult to compare data sets because they are derived under different conditions, presented in

different formats and benchmarked against different reference sets. This information must be accommodated in 'interoperable' repositories, and many databases are being set up for this purpose. This requires distributed databases from the outset that include common machine-readable tags for the species, strain or tissue used, say. By agreeing a minimal set of common standards, the proteomics community will ensure that their many databases can communicate with each other in the same language, and be queried as if they were a single resource.

In April this year, the Human Proteome Organisation founded the Proteomics Standards Initiative (<http://psidev.sourceforge.net>), which aims to define the standards required. This will be a huge challenge, not least because, as one computer scientist once quipped, "standards are so important that everyone wants their own". But if the full promise of proteomics is to be realized, the community and the bioinformaticians that service it need to pay more attention to the often thankless task of standard setting. Only then will it be possible to find order in the labyrinth of biomolecular networks. ■





Out in the cold
Frosty reception
for Bush's climate-
research plans
p595



All together now
European labs
network to track
transgenic foods
p596



Flexible focus
Australia unveils
its vision for
scientific research
p597



Shut down
Syngenta pulls
the plug on plant
genome centre
p599

NASA roadmap charts route to a quarter-century of exploration

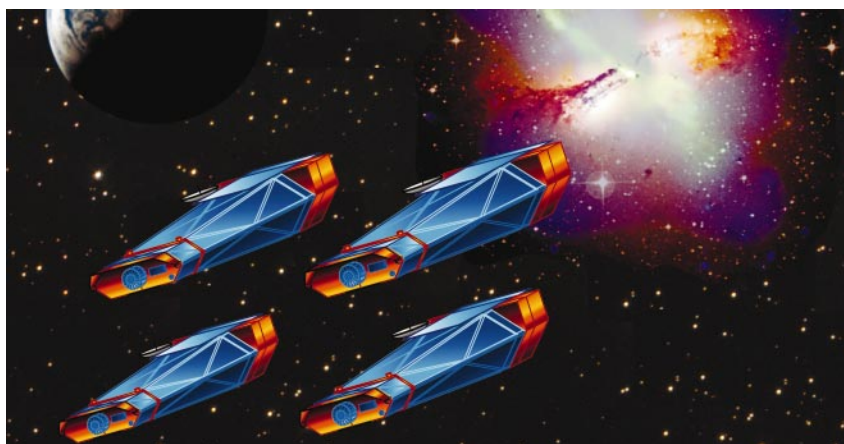
Tony Reichhardt, Washington

NASA is making preliminary plans for a multibillion-dollar programme to study black holes, dark matter and other cosmic exotica from spaceborne observatories.

Called 'Beyond Einstein', the initiative would be modelled on the agency's Origins programme, which focuses on galaxy and planet formation and the search for life in the Universe.

Like Origins and 'Living With a Star', which covers solar-terrestrial physics, Beyond Einstein is being designed as a coherent framework that would provide a rationale for the support of its scientific field over many years. It would include large space-based observatories, smaller 'Einstein Probes', costing up to \$500 million each, plus funding for further scientific research and technological development. The plan is set out in a soon-to-be-published 25-year strategic roadmap written by scientific advisers to NASA's existing Structure and Evolution of the Universe (SEU) programme.

NASA could request fresh funding for the initiative as early as the 2005 fiscal year, according to Anne Kinney, director of the agency's astronomy and physics division. She told an SEU advisory committee in



Search across space: Constellation-X will spearhead NASA's programme to study black holes.

Washington on 3 December that Beyond Einstein is NASA chief scientist Ed Weiler's highest-priority new initiative. If approved by the White House and Congress, the programme would probably be funded at a similar level to Origins — several hundred million dollars a year.

Beyond Einstein would build on the work of current spacecraft to address three main research questions: what powered the initial

inflation of the Universe following the Big Bang, what happens at the edge of a black hole, and what is dark energy?

The authors of the SEU roadmap, led by Sterl Phinney, a theoretical astrophysicist at the California Institute of Technology, considered two recent National Academy of Sciences reports in shaping their plan: a decadal review of priorities for astronomy missions, and another report on the interface between

Canada stops Harvard's oncomouse in its tracks

Erika Check, Washington

The Canadian Supreme Court has turned down a far-reaching patent application on a genetically modified mouse — setting Canada's intellectual property law on a strikingly independent path.

Harvard University had applied to patent the methods for making a transgenic mouse that is highly susceptible to cancer — the so-called 'oncomouse' — in Canada, the United States, Europe and Japan. The university had also asked to patent the actual mice, as well as all non-human mammals that have been modified to be vulnerable to cancer.

The patents have already been granted in all of these countries except Canada. But on

5 December, Canada's highest court denied Harvard's request to patent the mice. "A higher life form is not patentable because it is not a 'manufacture' or 'composition of matter,'" the court said in its ruling, which can't be appealed except through legislation.

The court did not rule out Harvard's claims on the process used to make the oncomouse, however, and the university says it will continue to pursue that claim.

The decision isn't seen as a huge setback for Harvard, because Canada is a relatively small part of the global market for the mouse. But it is expected to have major ramifications for life-sciences research and biotechnology in Canada. BIOTEC Canada, a

biotechnology association, says the move will deter researchers from trying to develop research tools, such as transgenic animals, that could be useful for science and medicine.

But environmental groups welcomed the decision. "We think the court got it right," says Jo Dufay, campaign director for Greenpeace Canada. "These issues are so complex that they require full public debate, and that goes beyond a simple tinkering with the Canadian patent act." The court ruling is expected to lend impetus to a major overhaul of Canadian patent law, which parliament is expected to take on before the current government leaves office.

► astronomy and physics (see *Nature* **416**, 775; 2002).

The roadmap's top two recommendations for 'Einstein Great Observatories' were also ranked highly in the decadal review. These are the Laser Interferometer Space Antenna (LISA), a European-US space-based array designed to detect gravitational waves, and Constellation-X, a cluster of X-ray telescopes that would conduct high-resolution spectroscopy of energetic objects, including black holes.

The smaller, more focused Einstein Probes would be proposed and managed by scientists outside NASA, and would be launched every three to four years. The first three candidate missions cited in the roadmap are a Dark Energy Probe, an Inflation Probe to study the remnant radiation from the Big Bang, and a Black Hole Finder Probe.

Agency scientists and roadmap team members have already begun briefing Washington policy-makers on the plan, which would need funding beyond NASA's current \$3.6-billion annual budget for space science. All are well aware of the need to sell the new initiative. Public communication is "absolutely critical to winning support for Beyond Einstein", Kinney told the advisory meeting.

But space scientists are confident that the programme can deliver both cutting-edge science and ample public enthusiasm. "These are sexy topics," says Rocky Kolb, an astrophysicist at Fermilab in Batavia, Illinois, who chairs the SEU subcommittee. ■

Paper trail reveals references go unread by citing authors

Philip Ball

Many of the references cited in scientific papers have not been read by the authors citing them, according to an analysis of how errors in citations propagate through the literature.

It isn't easy to establish directly — and truthfully — whether citations have been read. So Mikhail Simkin and Vwani Roychowdhury, two electrical engineers at the University of California, Los Angeles, decided to estimate author diligence by checking how often errors in citation lists are passed on through other papers. They reckon that scientists who have looked up and read a paper they want to cite are unlikely to misprint it in their own articles. The pair concluded that four out of five authors had not done their homework (M. V. Simkin and V. P. Roychowdhury Preprint cond-mat/0212043 <http://xxx.lanl.gov>; 2002).

Simkin and Roychowdhury tracked all of the citations of a seminal 1973 paper on condensed-matter physics (J. M. Kosterlitz and D. J. Thouless *J. Phys. C* **6**, 1181–1203; 1973). They looked at 4,300 citations of this paper and found that mis-citations of it were often identical to each other.

The probability of the same misprint occurring twice by chance is very small, so Simkin and Roychowdhury conclude that these errors are usually propagated by the

reference being copied from someone else's citation list. The most common misprint appeared 78 times.

Based on the number of distinct misprints, the two researchers estimate that only 22–23% of citations followed from a reading of the original paper. And they postulate that this is typical of the scientific literature as a whole.

Others agree that lazy citation is a real issue. Sidney Redner, a physicist at Boston University who has studied citation statistics, says that he thinks the latest estimate of original reading may be on the low side. "I might have guessed at about 50%," he says, but he agrees with the qualitative conclusion that many citations are unread.

How much does it matter? The Office of Research Integrity (ORI) at the US health department has recently drawn criticism for probing into this kind of practice, which is considered by some to be a kind of low-key misconduct (see *Nature* **420**, 253; 2002). If the problem is as widespread as the new study implies, it becomes harder to dismiss.

"I think it's very serious," says Kay Fields of the ORI, although she sees it as a matter of negligence rather than misconduct. Fields points out that failure to consult the original source can also propagate misconceptions about what the source's conclusions actually are. "It's poor research practice, for sure," she says. ■

Radar array tests the atmosphere at pole position

David Cyranoski, Tokyo

A three-metre radar antenna was among the equipment carried by Kaoru Sato as she set off for the South Pole late last month. Sato, who is based at the National Institute of Polar Research (NIPR) in Tokyo, will do a pilot assessment for a project that could make the Antarctic home to 1,000 such antennas set up in a circular array some 160 metres across.

The array would fill a major gap in our knowledge of the atmosphere, supporters of the project say. Already, radars around the world provide valuable data on phenomena such as atmospheric energy transfers. Researchers use the arrays to transmit radio waves. They then read their reflections to establish fluctuations in temperature, electron density and wind turbulence.

Such data offer information on how the different levels of Earth's atmosphere interact with each other and with the planet's climate — an important component in the study of global climate change and in

understanding global weather circulation. But according to Masaki Ejiri, head of the NIPR project, global models badly need more information about such patterns in the little-studied region of the South Pole's atmosphere. "The unknowns far surpass those anywhere else," he says of the area.

A better understanding of energy transfer in the atmosphere will also help to answer several other questions, Ejiri says, including whether the current improvement in the

ozone layer's condition is likely to be short-lived, and how aurorae affect temperatures in the middle atmosphere.

But Sato's bold mission won't be easy to complete. On current estimates, the circular array would consume more power than Japan's entire South Pole research base, says Toru Sato, Kaoru's husband and an engineer at Kyoto University. Sato believes he can reduce this energy consumption by half.

Other technical problems remain to be solved. One is ensuring that the equipment can withstand the low temperatures in the region. Another is interference with the radar signal by other radio waves. Sato's work with the single antenna will seek to measure this extraneous 'noise' with a view to figuring out how best to neutralize it.

But the project's biggest long-term challenge is to win the tens of billions of yen it would cost to build. Ejiri says that he is "convinced that the project is valuable enough to overcome these difficulties". ■



Japan's antenna project aims to shed light on the South Pole's atmosphere, including the aurorae.

Bush climate-change plan gets cool response

Kendall Powell, Washington

A plan proposed by the Bush administration to research climate change is as likely to obscure important questions on the subject as to answer them. Such was the prevailing view of scientists who met last week to discuss the project.

The meeting of 1,300 scientists and officials was held in Washington on 3–5 December to deliver feedback on the draft Climate Change Science Program, which was released by the Bush administration last month (see *Nature* 420, 110; 2002).

The plan will direct the US government's \$1.7-billion annual research programme on climate change, but its critics say that it lacks the clear priorities and detailed objectives needed to drive the programme forward.

Jim Anderson, an atmospheric scientist at Harvard University, said the draft was filled with "generalizations" that would serve the interests of "neither the public nor the scientific community".

Some climate scientists, including Anderson, hope that administration officials will take note of the feedback from the meeting so that the final version of the plan, due to be published next April, will be more specific. But others fear that the administration has no real interest in supporting research that might highlight either the causes or the dangers of global climate change.

Specialists at the meeting who had been

asked to review the plan said that its main elements — nine broad research programmes, plus some big initiatives such as a global climate observation system — might fail through a lack of clearly stated objectives.

The scientists also called for more research on the impact of climate change on specific regions of the United States, more study of the relationship between the water cycle and climate change, and more computer power for running climate models. But environmental groups dismissed the consultation exercise as window-dressing.

James Mahoney, director of the programme, which coordinates the research activities of 13 federal agencies, says that the final plan will accelerate federally sponsored research and focus it on key policy questions.

"It's difficult to provide political leaders with a mandate to give people medicine that will have major impacts — it's like asking, 'Who wants to raise taxes?'," says Mahoney, who is assistant secretary of commerce for oceans and atmosphere. He emphasizes the need to reduce the scientific uncertainties surrounding global warming so that policy-makers can make better decisions.

The plan would allow for an extra \$40-million annual initiative to tackle three areas of uncertainty: the role of aerosols in climate change, the size and location of carbon sources and sinks in North America, and how natural feedback processes influence climate change.



Clouded issue: can the United States' plan for climate research get to the heart of the problem?

NASA

But some observers question the plan's emphasis. "More research doesn't necessarily decrease uncertainty," says Benjamin Preston of the Pew Center on Global Climate Change. "And levels of uncertainty are not necessarily preventing decision-making right now."

Researchers are now hoping that Mahoney will produce a final version of the plan that is more to their liking.

♦ www.climatescience.gov

Safety panel backs principle of gene-therapy trials

Erika Check, Washington

Clinical trials of a gene therapy for a rare disorder of the immune system should be allowed to continue in the United States, according to the panel that advises the National Institutes of Health (NIH) on the safety of gene transfer.

Regulators in five countries halted such studies earlier this year, after a French boy being treated for severe combined immunodeficiency disease (SCID) developed a leukaemia-like illness (see *Nature* 420, 116–118; 2002). Most researchers now agree the gene therapy was a cause of his cancer.

The US Food and Drug Administration (FDA) has been working with investigators to determine how the country's own trials could proceed. Many researchers are now calling for extra precautions to check for cancer in certain patients undergoing gene therapy.

The NIH's Recombinant DNA Advisory Committee (RAC), which met in Bethesda, Maryland, on 4–6 December, recommended that the US trials should proceed with appropriate monitoring and informed

consent. But the RAC deferred any rulings on monitoring patients after such trials.

Scientists suspect that the retrovirus vector used to deliver the genetic material to the French patient caused his cancer. But the RAC did not comment on any US gene-therapy trials using retroviruses apart from those to treat SCID. It is still unclear how the FDA will deal with the other studies.

Current NIH guidelines do not require researchers conducting gene-therapy trials to follow their patients' health for any set period of time. But the FDA has already told investigators that they should make plans to follow their patients long after the studies are over. At least one scientist running a SCID gene-therapy trial told the RAC that he has been advised by the FDA to plan on tracking his patients for the rest of their lives.

RAC member David Sidransky, a cancer geneticist at Johns Hopkins University School of Medicine in Baltimore, says that it would be expensive and impractical to require more than a handful of investigators to undertake life-long monitoring. He adds that two of the

French boy's close relatives had childhood cancers, so he may have inherited a genetic defect that raises his risk of developing cancer.

"The question is would I rather spend \$100 million on monitoring, when we may not see another leukaemia in the next 100 years, or would I rather spend \$100 million treating sick patients," Sidransky says. "I'm not ready to commit to a major monitoring programme based on information from one patient."

Theodore Friedmann, a gene-therapy researcher at the University of California, San Diego, and chair of the RAC, says that the committee will take up the issue at its next meeting in March, and then draw up guidelines on how researchers should monitor their subjects. The FDA is expected to take the RAC's recommendations into consideration as it decides what to require from scientists who want to conduct gene-transfer studies in the future. "I hope that the language we come up with will be relevant not just to SCID, but to gene-transfer trials more generally," Friedmann says.

Labs cook up strategy to test transgenic food

Alison Abbott, Munich

The European Commission last week formally inaugurated a network of laboratories specifically charged with detecting genetically modified organisms (GMOs) in foods.

Officials hope that the labs will develop common standards for food testing, and will help to win public trust for genetically modified products by implementing new rules to govern the contents of such food.

As the network was launched, the agriculture ministers of the European Union (EU) were reaching agreement on what these regulations should say. But the rules will not be finalized until the European Parliament approves them, probably next year.

Given the public hostility to genetically modified food in Europe, most politicians support the explicit labelling of foods to inform consumers of their content.

The rules approved by the ministers would see food labelled if it comprises more than 0.9% GMOs — or products derived from GMOs — that have been approved by EU regulators. Products containing more than 0.5% GMOs that are thought to be safe, but have not yet been formally approved, would be banned from sale.

The European Network of GMO Laboratories, which employs about 450 people at more than 45 sites in the EU, will support the rules by validating and harmonizing methods for detecting specific GMOs in foods.

In the EU, about 80% of the products on sale, or awaiting approval, that contain GMOs



Serving suggestions: ministers hope to erode Europe's negative attitude to genetically modified foods.

have been validated on "individual initiatives by individual institutes", says Guy Van den Eede, a scientist at the Joint Research Centre's Institute for Health and Consumer Protection in Ispra, Italy, who chairs the network. "But it is important for everyone that these methods are validated in a consistent way across the EU." The network expects to validate between five and seven GMO tests next year, he says.

The labs will test only foodstuffs that contain DNA and proteins from genetically modified plants. The final EU regulations may also call for the labelling of highly refined foods made using genetically modified materials, but which no longer contain them. The food

industry has called this proposal unworkable.

The biotechnology industry has reacted positively to the laboratory initiative. "Industry has always stated that consistent methods for scientific traceability are necessary," says Simon Barber, a spokesman for EuropaBio, a group representing European biotechnology firms.

And many commission officials hope that having the labs in place will help to prepare for a lifting of the EU moratorium on new approvals for GMOs, which has been in place since 1998. So far, 18 such organisms have been approved, and 13 more are in the regulatory pipeline, pending the moratorium's end. ■

Syngenta ready to drop plans for Indian rice venture

K. S. Jayaraman, New Delhi

Syngenta, the Swiss-based agricultural biotechnology company, is set to withdraw from a collaborative deal that would have brought it commercial rights to a unique collection of Indian rice strains.

The ground-breaking deal would have given the company access to over 19,000 strains of local rice cultivars, painstakingly gathered by the agricultural scientist R. H. Richharia in the 1970s. In exchange, Syngenta would have provided support for collaborative research with the university that hold the seeds.

But plans to announce the agreement were postponed in October and, following harsh criticism from scientists, environmentalists and government officials, the company is now likely to abandon the plan, says Pawan Malik, president of Syngenta's seeds division in India.

Collaborations of this type are vital to the development of agricultural

biotechnology in the developing world — but concerns about ownership have made them very difficult to pull off.

Richharia's collection is held by the Indira Gandhi Agricultural University (IGAU) in Raipur, Chhattisgarh. The university and Syngenta were accused of excessive secrecy after the Chhattisgarh Biodiversity Security Forum, a non-profit group, revealed the existence of the negotiations. The IGAU's vice-chancellor V. K. Patil subsequently confirmed that the university had held three rounds of meetings with Syngenta since the summer to discuss an agreement whereby the two parties would jointly develop hybrid rice varieties.

Patil said that Syngenta would provide an undisclosed amount of research funding to the university in exchange for access to the collection, and would also pay seven years' worth of royalties on any new varieties it sold as a result. Patil said he

was going to make the deal public at the appropriate time.

But some scientists denounced the negotiations. "Richharia's collection is a national wealth and the IGAU has no right to treat it like its private property," says E. A. Siddiq, chairman of the research advisory committee of the National Bureau of Plant Genetic Resources.

The Indian Council of Agricultural Research, the main farm-research agency, has asked the university to clarify how it negotiated with Syngenta without its knowledge.

A spokesman for Syngenta says that the negotiations took place "to explore the possibility of working together to develop new rice hybrids that meet specific farmers' needs in that part of the country".

"We have 35 research collaborations in India, but this one has not worked out well," adds Malik. He declined to identify any of the others, however. ■

Australia sets priorities for future research

Carina Dennis, Sydney

All Australian government research agencies should be working towards environmental sustainability, promoting and maintaining health, developing new technologies and safeguarding the nation, according to a long-awaited list of priorities published by the government on 5 December.

Researchers have broadly welcomed the priority list — and the freedom that the government says it will allow agencies in sticking to it. “I’m pleased with the outcome — it will help to focus the mind but retain some flexibility,” says Kenneth Baldwin, a laser physicist at the Australian National University in Canberra, and chair of the policy committee of the Federation of Australian Scientific and Technological Societies.

Some had feared that the conservative government of Prime Minister John Howard would issue a far more prescriptive list of what Australian researchers should be doing. But agencies are expected to make marked changes to align their programmes with the new priorities — and may get more direct marching orders from the government if they don’t.

The exercise “sets a broad picture, and leaves a level of flexibility underneath,” says Vicki Sara, chief executive of the Australian Research Council (ARC), the nation’s main grant agency for non-biomedical research.

Some agencies have already initiated steps to redirect their programmes in anticipation of the government’s announcement. Earlier this year the Commonwealth Scientific and Industrial Research Organisation (CSIRO), which runs a network of public laboratories, said that it will redirect 30–40% of its budget into seven “flagship” programmes by 2006.

The community’s broadly positive response to the announcement was in marked contrast to the uproar over the ARC prioritization, announced in January, which was criticized for lack of consultation and for being announced too close to grant application deadlines (see *Nature* **415**, 724; 2002).

“Presumably, lessons were learned from that experience and consultation has been wider this time,” says Baldwin. For the December announcement, the government took advice from a panel of scientists and business leaders, held public discussions around the country, and read over 180 written submissions.

Australia’s research agencies are expected to outline plans to implement the new priorities within six months. “We will need to articulate our efforts if we’re to maintain a light-touch partnership with the government — if not, we could expect a more explicit directive,” says plant geneticist Jim Peacock, president of the Australian Acade-

my of Science and chair of the government’s advisory committee for priority setting.

The prioritization won’t be accompanied by any extra money, admits government chief scientist Robin Batterham. Nevertheless, he hopes it will help foster more research collaborations, and that these will be further encouraged by an audit of publicly funded research which the government plans to conduct in 2003.

“We expect to see new collaborations arising from this, as researchers identify common problems and goals,” says Peacock.

“There is a culture of independence across the agencies that needs to be overcome,” adds Graham Harris, an ecologist and chair of the CSIRO’s flagship programmes. Addressing that may prove tougher than it sounds, according to Sara: “There is a problem with fragmentation in Australian research,” she says, “and the challenge will be pulling people together.” ■



John Howard’s government will allow agencies to be flexible in following research guidelines.

D. LEWINS/AAP

Ousted creationist sues over website

Geoff Brumfiel, Washington

A Tennessee creationist is suing the operators of a popular physics website that refused to publish his alternative Big Bang hypothesis.

Robert Gentry, a lifelong Seventh-Day Adventist, filed the suit in the district court at Knoxville, Tennessee, against the operators of the arXiv preprint server, claiming that they refused a series of ten of his papers because of their religious content. Counsel representing the chief defendant, Cornell University in Ithaca, New York, says the claims have no merit and that the university has the right to choose what appears on the site.

Gentry, who has a masters degree in physics from the University of Florida, had papers in nuclear geophysics published in journals, including *Science* and *Nature*,

during the 1960s and 1970s. Those papers, he says, inspired him to come up with an alternative Big Bang hypothesis, which he submitted unsuccessfully to academic journals. He then tried posting his articles on the arXiv preprint server — a non-peer-reviewed website where physicists often post papers before submitting them to journals. When arXiv curators removed the papers and revoked his posting rights in 2001, Gentry complained, then filed the suit to regain access this August. “I’m a creationist and a believer in the Bible, but I want to know the truth. I want these papers to be tested by the scientific community,” he says.

Paul Ginsparg, a professor at Cornell and creator of the site, declined to comment, citing the ongoing suit. But Nelson Roth, Cornell’s associate counsel in charge of litigation, says that the rejection was based on Gentry’s lack of academic affiliation, not his beliefs. “The religious views of the plaintiff are completely irrelevant,” he says.

Even if the legal case makes no progress, it highlights some problems associated with websites whose content is not peer-reviewed, says Adrian Melott, a cosmologist at the University of Kansas in Lawrence. Melott, a co-founder of Kansas Citizens for Science, a group that has successfully lobbied against teaching creationism in the state’s schools, says he’s noticed a rise in “flaky” publications on the section of the arXiv server that he uses most. “We’re coming to a crunch” over what can be published on open servers, he says. ■



Europe's astronomers reflect on plans for giant telescope

London Astronomers from across Europe are joining forces to design a ground-based telescope with the largest mirror that is technically feasible. They will announce their ambitious plans at the Royal Astronomical Society meeting in London on 13 December.

The astronomers are embarking on a detailed design phase of the European Extremely Large Telescope (ELT). The telescope will have a mirror diameter of more than 30 metres, perhaps even three times that, says Gerry Gilmore, an astronomer at the University of Cambridge, UK, who chairs the ELT steering committee. A mirror more than 60 metres across would be needed to study some celestial objects, such as planets around other stars, or the oldest stars in other galaxies.

The annual running costs of such a telescope, which could cost up to £1 billion (US\$1 billion) to build, might be more than today's entire budget for astronomy in Europe, warns Gilmore. There are competing projects in the United States, such as the Giant Segmented Mirror Telescope at the National Optical Astronomy Observatory in Tucson, Arizona, and the California

Extremely Large Telescope, planned by the California Institute of Technology and the University of California.

Submarine view sinks hopes of end to *Prestige* oil spill

Madrid Oil is leaking from the sunken wreck of the oil tanker *Prestige*, Spanish officials said last week. Pictures taken from the *Nautille*, a French research submarine, show that strands of nearly solid fuel oil are trickling from several cracks in the bow of the stricken tanker. Marine scientists had predicted that the cold temperatures on the seabed would freeze the oil and prevent it



Oil leaking from the sunken *Prestige* fuels fears of further slicks.

escaping (see *Nature* 420, 347; 2002).

The vessel sank in over 3 kilometres of water some 250 kilometres off the coast of Spain on 19 November, taking about 60,000 tonnes of oil down with it. No sign of the solid fuel had been observed on the surface as *Nature* went to press. Between 10,000 and 20,000 tonnes of oil has

already spilled into the sea, contaminating approximately 400 kilometres of coastline.

South Korea talks of joining fusion project

Washington South Korea is the latest country to declare itself a potential participant in the ITER project, an international effort to build a US\$5-billion experimental fusion reactor.

Korean officials are in Barcelona this week to meet representatives from the project's principal partners — the European Union, Japan and Russia — to discuss the level of the country's participation. South Korea is already building a small but advanced fusion reactor, known as the Korea Superconducting Tokamak Advanced Research (KSTAR) project, in collaboration with the United States. ITER's partners held a similar meeting in Tokyo in November with officials from China, which is also considering signing up to the plan (see *Nature* 419, 545; 2002).

Meanwhile, a team from the US Department of Energy has completed a review of ITER's cost estimate and describes it as "generally reasonable". The United States pulled out of ITER in 1999, but senior officials said in the summer that they are considering rejoining the project.

♦ www.iter.org

Britain drops initiative to explain science to the public

London The plug has finally been pulled on one of Britain's first efforts at helping the public get to grips with science.

On 9 December, the Royal Society, the Royal Institution of Great Britain and the British Association, the three original funding organizations of the science-communication body Copus, announced that they are abandoning the search for a new leader and disbanding its ruling council. The move follows attempts to modernize the organization and the recent angry resignation of its previous chair, Bridget Ogilvie (see *Nature* 417, 577; 2002).

In a joint statement, the three organizations said: "We have reached the conclusion that the top-down approach which Copus currently exemplifies is no longer appropriate to the wider agenda that the science communication community is now addressing."

Australia opens way to embryonic stem cells

Sydney Months of heated debate in the Australian Senate about embryonic stem cells drew to a close on 5 December, when politicians voted to allow researchers access

to surplus human embryos created by *in vitro* fertilization. Researchers will now be able to derive new stem-cell lines from some 70,000 frozen human embryos created before 5 April, the date of a previous government agreement on embryo research. Both therapeutic and reproductive cloning have already been outlawed in a separate bill.

As in other countries, the issue has divided the government. It also attracted plenty of media attention, including a row over whether Alan Trounson, a stem-cell researcher at Monash University in Melbourne, deliberately misled the public when discussing stem-cell research (see *Nature* 419, 4; 2002). The National Health and Medical Research Council will design a regulatory framework and researcher-licensing system to cover the use of embryos.

Syngenta takes its rice genome institute off the boil

San Diego Agribusiness company Syngenta is to close its Torrey Mesa Research Institute (TMRI) in San Diego, California, with the loss of around 80 jobs, including scientists and support staff. TMRI is known for having helped produce a commercial version of the rice genome sequence, which was published alongside a publicly funded version this April (see *Science* 296, 92–100; 2002).



Field work: the rice genome was Torrey Mesa's last success.

About 75 TMRI scientists will move to nearby biotechnology company Diversa, which has signed a seven-year, US\$118 million contract with Syngenta to collaborate on plant-science work. The remaining 35 researchers will move to other Syngenta laboratories at

Research Triangle Park in North Carolina.

TMRI president Steve Briggs will switch to Diversa, which specializes in identifying microbes from extreme environments. Stephen Goff, who oversaw Syngenta's role in the rice genome-sequencing project, will move to Research Triangle Park.

Clarification The Climate Action Network Europe (CAN Europe) study of hydrofluorocarbon emissions, reported in *News (Nature* 419, 656; 2002), contained an error. The emission of these gases is predicted by CAN Europe to be much lower than the number given in its report at the time *Nature's* article was written. The report has now been corrected; see www.climnet.org/pubs/ozoneclimate.htm for the correct version and further details of the calculation.

Into unknown territory

The Alliance for Cellular Signaling is exploring new frontiers, both in fundamental scientific terms and in the way in which research in cell biology is conducted. Alison Abbott reports.

“We are exploring a vast, vast, vast continent,” pronounces cell-signalling guru Henry Bourne of the University of California, San Francisco, “and we know only a few ports, a handful of rivers and a couple of mountain ranges.”

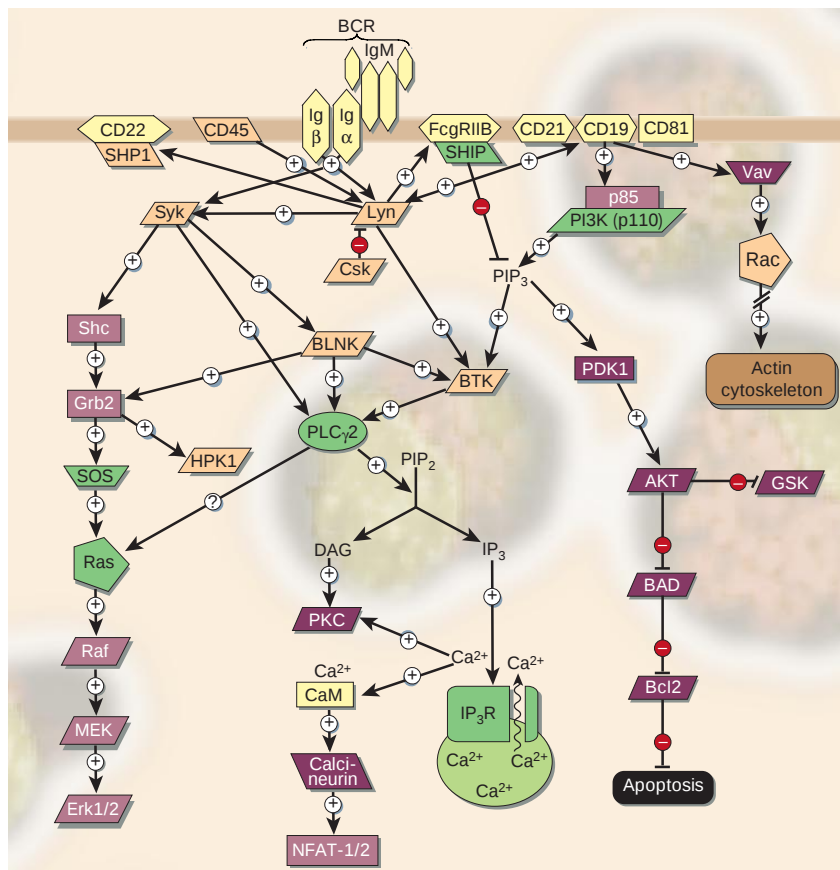
The dark continent that Bourne refers to is the labyrinth of overlapping biochemical signalling networks through which our cells make sense of the diverse internal and external signals that constantly bombard them. He is one of the founders of the Alliance for Cellular Signaling (AfCS), a ‘big biology’ project established to explore and colonize this territory.

This week, the project’s leaders plant a flag in a newly constructed frontier fort. Together with *Nature*, the AfCS is launching the Signaling Gateway, an online resource that will combine news and reviews with scientific databases. The project’s goals are also discussed in a series of articles in this issue (see page 703).

The intended beneficiary of the AfCS’s colonial venture is the wider community of cell and molecular biologists. As yet, few outside the project are aware that it is already generating data and making them freely available — or that the AfCS is soon to offer a range of research materials, including complementary DNA (cDNA) clones representing genes for signalling proteins.

So what is the AfCS, and what are its aims? The brainchild of Al Gilman of the University of Texas Southwestern Medical Center in Dallas, winner of a Nobel prize in 1994 for his discovery of G proteins and identification of their role in cell signalling, the AfCS has the grand vision of creating a ‘virtual cell’ — a complete model of its biochemical behaviour. By mapping signalling networks, it should theoretically be possible to predict exactly how cells will respond to particular combinations of signals. It sounds simple enough, but the task is almost infinitely complex, and some experts predict that it may defeat even the grand alliance of specialists that Gilman has assembled.

Cell biologists now realize that they have oversimplified the biochemistry of signalling pathways. They have long hypothesized that a signal, such as a hormone binding to its receptor, initiates a wave of excitation through a simple, linear chain of proteins. But it has become clear that the proteins don’t work individually, but instead assemble fleetingly into molecular machines, the components of



Protein puzzle: this simplified flurry of signals is just one of many hundreds of interacting pathways in mouse B cells (pictured in the background), highlighting the sheer scale of the alliance’s endeavour.

which are shared between pathways (see *Nature* **417**, 894–896; 2002). “The inside of a cell is a thick soup of proteins talking to each other in ways we just don’t understand,” observes Robin Irvine, a pharmacologist at the University of Cambridge, UK.

Signal strength

Gilman and his colleagues have produced a two-pronged plan for the AfCS’s immediate future. First, they will collate existing information about components of signalling pathways, standardizing its format in a database of ‘Molecule Pages’ that will present key data on thousands of proteins involved in cell signalling. Second, they will generate robust new data about pathways in two types of mouse cell — B lymphocytes and cardiac muscle cells — in a consistent way. The

project’s funding comes from one of the ‘Glue Grants’ awarded for multi-site projects by the National Institute of General Medical Sciences in Bethesda, Maryland, expected to total \$25 million by 2005, and from pharmaceutical companies. The latter have all signed up to the AfCS’s core philosophy that all data, plus as many research tools as possible, should be made immediately and freely available to all scientists.

Early versions of the Molecule Pages are now available on the Signaling Gateway. There is a page for each signalling protein that has been identified in the mouse — currently more than 3,000 — containing information imported automatically from gene and protein databases such as GenBank and SWISS-PROT. Each page includes information such as the protein’s molecular mass and isoelectric

lipid compounds at a time — is a new field, he says, “so I’m just grateful to have the opportunity to develop the technology”.

Some AfCS labs are dedicated to the development of molecular tools, such as double-stranded RNAs for use in RNA-interference gene-silencing experiments, and cDNA clones. As soon as they have been fully catalogued, such tools will be made available at cost price to the scientific community through the American Type Culture Collection in Manassas, Virginia.

Highly specific antibodies targeted at signalling proteins, meanwhile, are being developed at the AfCS Antibody Laboratory at the Southwestern Medical Center, in collaboration with companies that will manufacture and sell them. A current priority is the development of antibodies that can distinguish between active and inactive signalling proteins, which often differ only in the presence of a single phosphate group.

“Access to better reagents will be important, since this is what is most limiting to research at the moment — many reagents that you buy just don’t work,” says Julian Downward, who heads the signal transduction laboratory at Cancer Research UK’s London Research Institute.

Like all large projects, the AfCS has met its share of obstacles. Database interfaces designed to make it easy for authors to feed information to the Molecule Pages, and for users to quiz the data, are not yet ready. “The bioinformaticians in the collaboration were far too optimistic,” says Gilman, who adds that bioinformatics will have a greater share of resources in future. The bioinformatics bottleneck has delayed the cataloguing of molecular tools, which is postponing their availability.

But despite such setbacks, those who are familiar with the AfCS are optimistic about its prospects. “The alliance will eventually develop a new language to allow us to describe the large networks of signalling pathways whose complexity already has us throwing up our hands in despair,” says Michael Berridge, head of the signalling programme at the Babraham Institute in Cambridge, UK, and one of the few Europeans who sit on an AfCS advisory committee.

For now, producing a complete model of cell signalling is a distant dream. And the project’s leaders accept that there is a chance that gathering data on such a massive scale will not yield advances of the interest and importance that they hope for. “We believe intellectually that this is the right thing to do, but there can be no guarantees,” says Gilman.

“The collaboration itself is the biggest experiment of all,” suggests Bourne. “After all, the scientific culture of biology is traditionally very individualistic, and it will be interesting to see if scientists can work as a large and complex exploratory expedition.” ■

Alison Abbott is *Nature’s* senior European correspondent.
 ▶ www.signaling-gateway.org

Rally the troops: Al Gilman (foreground) has assembled a crack squadron of cell-signalling specialists, marshalled by teleconferencing. The team aims to generate data on signalling in two model cell types.

point — the pH at which its net charge is zero — as well as its sequence, and the sequences of its counterparts in other species.

Within the next few months, when all of the interfaces are ready, it will be possible to interrogate the data. If you know some of the characteristics of an unknown protein, the Molecule Pages database will be able to deliver a list of all the possible candidates. “It’s a very good place to start when you need a quick entry into the properties of a protein you have just come across,” says Bob Michell, a cell-signalling biochemist at the University of Birmingham, UK, who is not involved in the AfCS.

Over the next few months, these basic data will be supplemented by around 1,000 experts, each of whom will author one or a few Molecule Pages, adding information on each protein, including its functions and known interactions. Peer review will be organized by *Nature’s* editors in London, independently of the AfCS. In addition, associate editors drawn from the research community will help to oversee dozens of pages in their particular areas of expertise.

“Eventually, the Molecule Pages will be important in putting together the whole picture of how a cell functions — something we can’t even conceptualize yet,” says Irvine, associate editor for some 80 pages on signalling pathways that depend on molecules called inositol lipids. “But in the meantime they will be a remarkable source of information.”

Being an author of one of the Molecule Pages, which will be updated annually, will be time-consuming. So Gilman says it is essential that the effort should be recognized by faculty committees and granting bodies, in much the same way that they consider the value of authoring a widely cited review article. “If large-scale collaborations in biology are going to work, the community will have to

change its ways of evaluating effort,” he says.

In parallel to developing the Molecule Pages, the alliance has set up eight US laboratories, with almost 40 scientific staff, to generate signalling information on the two types of cell chosen as models. A further 40 or so scientists serve on committees that direct this research through teleconferencing.

The labs are now beginning to generate data that are being placed on the web. These include a screen for protein–protein interactions involving, so far, 30 signalling proteins in B lymphocytes, conducted in collaboration with Myriad Genetics of Salt Lake City. Another study is using a variety of techniques to study changes in gene expression, lipid biochemistry and other parameters in B lymphocytes exposed to 32 different ligands — small molecules that activate the cells. Eventually, the data being generated by the labs will be integrated into the Molecule Pages.

Admiring the view

“The collection of so much data on two model cell types is the most innovative part of the project,” says Walther Rosenthal, head of the Research Institute for Molecular Pharmacology in Berlin, who is not part of the AfCS. “Nowhere else can you find this integrative view of the cell.”

Anyone may publish analyses of the AfCS’s data in the usual way — but the project’s own scientists must wait a month after the data are posted on the web before submitting a paper to a journal. “This means that the alliance’s scientists have no special privileges,” says Gilman. For Alex Brown, who runs the AfCS lipidomics laboratory at Vanderbilt University in Nashville, Tennessee, the controls on publication are not important. Lipidomics — the high-throughput analysis of large numbers of

Share and share alike?

Producing a popular research tool will make you a lot of friends in science. But meeting requests to supply the material puts a heavy burden on your lab. David Cyranoski examines a system under pressure.

Yoshihide Hayashizaki is exasperated. Last year his team published a paper¹ describing a collection of mouse complementary DNAs (cDNAs) representing the coding sequences of thousands of genes. Now he receives a request to supply these key research tools, on average, almost every day. A new article² expanding the collection to some 61,000 cDNAs, published last week, is likely to increase that demand. But in trying to meet these requests, he has stumbled into a logistical, financial and legal minefield. "It's very stressful," says Hayashizaki, who is based at the Genomic Sciences Center in Yokohama, part of RIKEN, Japan's Institute of Physical and Chemical Research.

Hayashizaki's case is extreme. But researchers around the world are facing increasingly burdensome requests for a range of research materials including cDNA clones, antibodies and knockout mice. And they are expected to fulfil them. Many journals, including *Nature*, require that materials used in the papers that they publish are made available to other researchers. This allows experiments to be repeated, and lets scientists build on each other's achievements rather than starting from square one.

Until recently, mutual goodwill ensured that this almost always happened. But supplying materials can be expensive. The 21,000 cDNAs described in Hayashizaki's first paper cost more than US\$10,000 to prepare and ship. Journals do not require researchers to make materials available free of charge, but it is unclear how deeply suppliers are expected to dig into their own pockets to cover administrative and other costs. Those involved must also grapple with legal issues surrounding the transfer of materials. And as projects get bigger, materials more expensive and



Stressed: for Yoshihide Hayashizaki, sharing his mouse cDNAs has proved a logistical nightmare.

commercial interests more dominant, many researchers are asking how long the current system can continue to function.

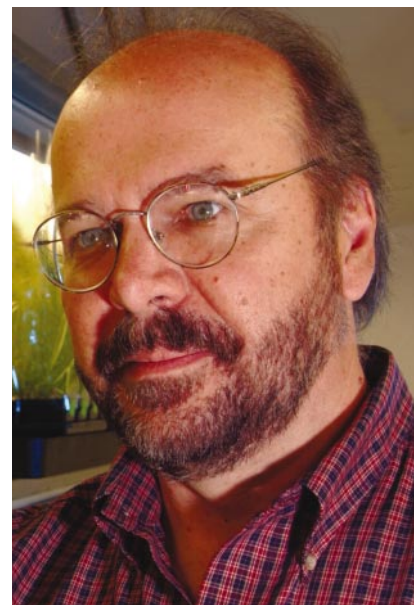
Distributing materials is time consuming as well as expensive. Collections must be maintained and orders processed. Shipping itself varies from the trivial task of popping plant seeds into an envelope, to the more intensive chore of putting antibodies on dry ice or express-mailing live animals. Ensuring that the proper materials are sent, and then recouping the costs, can be a full-time job for some technicians or research assistants. "It is an enviable problem, as it means you are getting many citations and getting your name established," says Thomas Cech, president of the Howard Hughes Medical Institute (HHMI) in Chevy Chase, Maryland, the

largest biomedical research charity in the United States. "But it can really add a lot of time and financial pressure."

For those working with widely used organisms, such as the fruitfly *Drosophila melanogaster* or the thale cress *Arabidopsis thaliana*, shifting the responsibility to public repositories is often the best solution. Joe Ecker works on *Arabidopsis* at the Salk Institute for Biological Studies in La Jolla, California. He has produced seeds of many mutant varieties, as well as tools such as bacterial artificial chromosomes (BACs), which can be used to clone sections of DNA. "The requests would have been a major distraction," says Ecker. "I did not want to turn my group into a stock distribution centre." So Ecker sent his materials to the *Arabidopsis* Biological Resource Center (ABRC), a collection and distribution service based at Ohio State University in Columbus. The stocks he sent more than doubled the ABRC's holdings, but the extra user fees they generated have so far covered the costs involved.

All would be fine if there were repositories for all areas of science, with each having the money needed to fulfil requests. But BACs and other genomics tools are being created in increasing numbers, and money and space at the repositories is already tight. Mutant-mouse storage houses, for example, are filling up as fast as they are built³.

Contracting the job out to the private sector is an alternative for some. Siamon



Taking stock: Joe Ecker (above) solved his distribution problems through the *Arabidopsis* Biological Resource Center (left).

Gordon of the University of Oxford, UK, found that the best way to handle his popular monoclonal antibodies, which have potential therapeutic applications as well as uses in academic research, was to license them to the nearby reagent supplier Serotec, which distributes them for profit. "The arrangement has worked very well," he says. Gordon will also provide materials at his own expense for researchers who cannot afford them. "This does not get specific grant support, but hopefully gets us brownie points," he says.

Supply and demand

But this option has its drawbacks. Reliance on commercial distributors jacks up the cost of research, limiting the potential pool of users. And it isn't always a viable option, as only the most popular and profitable materials will attract the interest of a distribution company. Commercial suppliers have their place, say scientists, but the funding of public repositories is vital if research tools are to be made as widely available as possible. "We need continued support to allow the distribution of stocks that are important but not necessarily profitable," says Randy Scholl, director of the ABRC.

Even when public or private systems do cover distribution needs, researchers can find themselves thumbing through legal documents that need a lawyer to evaluate. Many researchers, for instance, have complained about the lengthy forms Hayashizaki's insti-

tute requires recipients to sign before it will release the clones. Such material transfer agreements (MTAs) are used for a variety of reasons. They can, for example, restrict the ways in which the recipient can use the materials, especially by forbidding potentially dangerous uses such as clinical experiments.

For the sender, MTAs have obvious benefits. They can ensure that the materials do not get degraded through copying, for instance. Ecker says that he once used an MTA to distribute yeast artificial chromosomes, another tool used for cloning DNA. "We simply were saying, 'If you want the clones, get them from us,'" he says. "Not all MTAs for genomic resources are a bad thing."

But MTAs sometimes contain restrictions relating to publication and property rights that are unacceptable to would-be recipients or their employers. Some agreements, for example, require that the sender receives a part of the profits from any commercialization of the research, or that any potential publication must be authorized by the sender. "For almost every MTA, there are some researchers whose institution will not allow them to sign it and thus receive the associated material," says Scholl.

Thankfully, researchers are trying to minimize the use of such agreements. "We try to ignore MTAs when sending," says Claude Desplan, a developmental biologist at New York University, who supplies mutant lines of *Drosophila*. When sending his seed stocks and

BACs to the ABRC, Ecker says he convinced administrators at the Salk Institute that his project would have "the greatest impact on the plant-biology community with no strings attached", and so was able to avoid using an MTA. Cech adds that the HHMI's form is only about three sentences long.

Repositories, especially in the United States, are also taking action, putting pressure on collections elsewhere to follow suit. The Mammalian Gene Collection, a set of cDNAs supplied by the National Institutes of Health (NIH) in Bethesda, Maryland, does not use MTAs, and the ABRC this year decided to refuse any further donations with MTAs attached. "This will place the burden of distribution of these large and unwieldy collections squarely on the shoulders of the individual investigators who have produced them," says Ecker. "This could force institutions to reduce restrictions on genomic-scale materials."

Ripped red tape

Indeed, some repositories have found themselves forced to drop MTAs. In November 2001, the RZPD, Germany's human-genome resource centre in Berlin, which recently began distributing a valuable library of human cDNAs, shifted from a lengthy MTA to a brief "good faith agreement". This merely frees the RZPD from responsibility for any problems with the material itself or dangerous uses of it. The move was prompted by the realization that its old MTA was forcing researchers in Germany to obtain resources from institutions in the United States.

Even if researchers can get around MTAs, other legal problems sometimes await — as Hayashizaki has found out. Mouse cDNAs are created from the messenger RNA (mRNA)

molecules that are transcribed from active genes. The mRNAs are mixed with an enzyme called a reverse transcriptase, which latches on to the mRNA and copies it into the more stable and easily manipulated cDNA. Hayashizaki used one of the most popular reverse transcriptases available: SuperScript, produced by Invitrogen of Carlsbad, California.

Hayashizaki realized that distributing the clones would place an intolerable burden on his lab. So in 2000, he licensed them to Dnaform, a Tokyo-based biotechnology company spun off from RIKEN. The firm planned to distribute the set to academic researchers at cost (US\$12,000) and to commercial organizations for US\$250,000. SuperScript is not present in the final cDNAs, so Hayashizaki assumed that Invitrogen had no claim over Dnaform's materials. But he had missed the small print, in which Invitrogen claims that its US patent covers any cDNAs made with the enzyme.

Early last year, Dnaform received a letter warning that it had no right to distribute its cDNAs for profit in the United States. Invitrogen also argues that the 'at cost' academic agreement blurs the line between profit and non-profit, as each cDNA comes in at roughly three times the cost of those from the American Type Culture Collection, a non-profit bioresources company in Manassas, Virginia. But Dnaform president Toshizo Hayashi responds that the fee still doesn't fully cover the cost of the high-quality clones that RIKEN and Dnaform provide.

Catch 22

The dispute between Dnaform and Invitrogen leaves Hayashizaki's RIKEN team stuck with the responsibility of dealing with distribution. "In Japan, the patent office would never accept this broad coverage," Hayashizaki says. "If the United States does, then it makes sense for Invitrogen to pursue its rights. But I think it is unreasonable."

Invitrogen's patent is currently being challenged in US courts by other firms wanting to sell reverse transcriptases. But in the meantime, the only way for a US group to get cDNAs from the first collection is through a purely academic collaboration with Hayashizaki's lab, of which 330 have already been set up. Hayashizaki has tried to find a



Difficult times: Thomas Cech believes problems with materials sharing are getting worse.

way around using SuperScript, but fears that his second clone set might also get caught up in the legal morass.

To avoid similar problems in the future, researchers will have to think carefully about the intellectual property rights associated with the reagents they use. But is everyone likely to put his or her heart into this? Disturbingly, many researchers suspect that legal problems, obstructive MTAs, or logistical difficulties are sometimes used as a smokescreen by researchers who simply don't want to share.

In January, a team led by Eric Campbell of the Institute for Health Policy at Massachusetts General Hospital in Boston reported that almost half of 1,240 human geneticists they surveyed said that they had been refused access to data, information or materials in the past three years. Problems with accessing materials accounted for more than a third of all refusals⁴.

Researchers often cite difficulties with MTAs as a reason for not sending out materials, but four-fifths of those in the survey who admitted to not honouring a request said that they had been put off by the time and effort involved. And some researchers feel that, in a minority of cases, the real reason is a desire to maintain a competitive advantage over peers who lack access to key tools. "Some investigators clearly hide behind their institution's restrictive MTAs," says Ecker. "They don't really want to make their published materials freely available."

Cech believes this is especially true in industrial research centres, where researchers want to hold on to their materials to make profit from them. He claims many companies will offer to supply published materials with an absurdly restrictive MTA. Would-be recipients bring the MTA back to their institution's lawyers only to find out that it is not possible to sign the agreement. A lengthy back and forth

negotiation then ensues. "The idea is to be as obstructive as possible," says Cech.

Researchers who are just starting their laboratory may be hardest hit, especially when it comes to asking for materials from foreign colleagues — who know that they are unlikely to be penalized for being obstructive to a junior researcher in another country.

Three years ago, Ray Truant moved from a postdoctoral position at Duke University in Durham, North Carolina, to open a Huntington's disease research laboratory at McMaster University in Ontario, Canada. He says that one US researcher ignored 12 e-mails, four phone calls and several letters before finally reneging on a promise made in person to supply plasmids, loops of bacterial DNA that can be used to clone genes. "As we are not in their granting systems, we are unlikely to review their grants and therefore there is no reason to be courteous," laments Truant.

Easy access?

Cech believes that problems with the sharing of research materials have got worse over the past 10 years. So is it possible to differentiate those who are truly too burdened by paperwork and administrative costs from those who merely do not want to do it? And how can those involved enable the former, and force the latter, to send materials? The US National Research Council is currently sponsoring an investigation into community standards for sharing publication-related data and materials, which may provide some suggestions. The committee, chaired by Cech, is expected to release its report soon.

Cech is reluctant to comment on what ideas the report might contain, but those who have sent and received materials say that everyone involved can play their part. Repositories, for example, can help by refusing to accept materials that come with MTAs, or by only accepting simple and non-restrictive MTAs. Policing by journals could also be part of the solution. Editors could encourage researchers who run into problems to get in touch, for instance, and then pursue miscreants. Funding agencies are an additional piece of the jigsaw. Cech says that the HHMI is amenable to providing extra funding to help the developers of research materials to distribute them to others. But many other agencies, including the NIH, provide no such grants.

New grants to help sharing, together with pressure from repositories and journals, would almost certainly improve the situation. Despite his unfortunate experience, Truant believes that a combination of carrot and stick could help to ensure that standards of behaviour aren't allowed to slip. ■

David Cyranoski is Nature's Asian-Pacific correspondent.

1. The RIKEN Genome Exploration Research Group Phase II Team and the FANTOM Consortium *Nature* **409**, 685–690 (2001).
2. The FANTOM Consortium and the RIKEN Genome Exploration Research Group Phase I & II Team *Nature* **420**, 563–573 (2002).
3. Abbott, A. & Knight, J. *Nature* **417**, 785–786 (2002).
4. Campbell, E. G. et al. *J. Am. Med. Assoc.* **287**, 473–480 (2002).



Material gains: California-based Invitrogen claims rights over cDNA clones made using its enzymes.

Isolationism is not the answer to bioterrorism

Increased support for research in the developing world would be a better strategy.

Sir—It has been suggested that the best way for the United States to keep biotechnology-enhanced biological weapons away from terrorist groups is for it to support related research and training only for US scientists (see, for example, *Nature* **414**, 3–4; 2001). We do not believe such an isolationist attitude to be useful. Increasing research support and training in biotechnology and genomics to scientists in the developing world may be the best way to prevent bioweapon attacks on the United States and its allies.

Inadequate funding by northern scientific bodies — of other scientific disciplines as well as genomics and biotechnology — marginalizes southern scientists and widens the gap between northern and southern scientists' mindsets. If southern scientists are given a stake in the northern system, through sponsored research opportunities, its scientists will be less likely to want to help terrorist actions against northern interests.

Southern scientists are, of course,

making significant novel contributions to science, which could include assisting the development of biodefence strategies. Their contributions are not only welcome, but enhanced northern research support could also challenge the stereotype that the north does not have the interests of the south at heart — a view propagated by anti-northern extremist groups. Scientists everywhere, south or north, need to be aware of the regulatory and ethical implications of bioweapon proliferation. Sponsored training by northern funding agencies is the best way to achieve this end.

Such training in genomics and biotechnology could be made conditional upon such scientists passing stringent intelligence review and verification by the sponsoring institution or its authorized representative, and their states being signatories to chemical- and biological-warfare conventions and protocols. Given the potential trade and investment opportunities that come with a skilled, biotechnologically competent workforce,

sponsoring advanced training of southern scientists in genomics or biotechnology could serve as an incentive for countries to sign and comply with the conventions.

The Fogarty International Center of the US National Institutes of Health (NIH) has budgeted approximately \$45 million for the training of international scientists for fiscal year 2002–03. Although this approach is commendable, apart from the dilution of the money over all the sub-disciplines in health, the amount is minuscule compared with the \$1.75 billion the US government has allocated to the NIH for biodefence research alone in 2003. Increasing research support and training in biotechnology and genomics would make it easier for northern scientists to say to their southern colleagues: "We're acting with you, so don't act against us."

Jerome A. Singh*†, **Peter A. Singer†**

*Howard College School of Law, University of Natal, Durban, 4041, South Africa

†University of Toronto Joint Centre for Bioethics, 88 College Street, Toronto M5G 1L4, Canada

Schools can be inspired by a summer of science

Sir—We appreciated reading your recommendation that other universities should promote initiatives like those offered in the United States, to "give young people enough of a glimpse of the world of science to be enticed further into it". In particular, we liked your suggestion that research facilities should invite secondary-school students to their laboratories during the summer holidays (see *Nature* **419**, 233; 2002).

In Portugal, research institutes have been providing summer internships for secondary-school students for several years now, with the support of European Regional Development Funds and the Portuguese government (see www.cienciaviva.pt). Students work in research laboratories, where they are given specific simple scientific tasks, sometimes including field work, for one week or more. The aim is to give them a clear idea of the realities of research in topics including mathematics, biotechnology, robotics and cancer. Several of the best Portuguese scientific institutes participate in this activity on a regular basis, and both students and researchers have found it to be a very positive experience.

Last summer, a similar initiative was extended to secondary-school teachers, who are given a chance to work with state-

of-the-art laboratory equipment and learn about the most recent developments in scientific and technological research.

Rosalia Vargas,
Ana Noronha

Ciência Viva — National Agency for Scientific and Technological Culture, Pavilion of Knowledge, Parque das Nações, Alameda dos Oceanos, 1990-223 Lisboa, Portugal

What has posterity done for us? It's not the point

Sir—To understand why many physical scientists regard economists with scepticism, one need look no further than the Concepts essay on discounting ("An eye on the future" *Nature* **419**, 673–674; 2002) by L. H. Goulder and R. N. Stavins. After describing an example in which a \$4-billion investment now would prevent us from causing \$800 billion of environmental damage 100 years hence, they ask: "If future generations do not actually compensate the present one, is it still appropriate to enact the policy?"

Who but an economist could imagine that future generations would owe us an impossible debt for not damaging their environment? Isn't it we who owe future generations a sound environment? Discounting provides a well-defined measure relating present and future sums.

The problem is that this measure is not particularly useful for problems involving intergenerational transfer.

Ken Caldeira

Climate and Carbon Cycle Group, Energy and Environment Directorate, Lawrence Livermore National Laboratory, 7000 East Avenue, L-103, Livermore, California 94550, USA

Goulder and Stavins reply—We agree with Caldeira's view that it would be unfair to require future generations to pay the present generation for the costs of current climate policy. Our example was meant to illustrate complications that arise in evaluating policy options involving winners and losers.

Caldeira goes on to claim that discounting (and benefit–cost analysis) is not useful for problems with intergenerational impacts. Here we differ. The aggregate policy-generated gains and losses, translated by discounting into comparable units, are highly relevant for assessing public policies. But we emphasize here, as in our essay, that other considerations — including attention to other criteria of fairness — are important in policy evaluation.

Lawrence H. Goulder*,
Robert N. Stavins†

*Department of Economics, Stanford University, Stanford, California 94305-6072

†John F. Kennedy School of Government, Harvard University, Cambridge, Massachusetts 02138, USA

The conscience of physics

The life, work and dreams of Wolfgang Pauli.

No Time to be Brief: A Scientific Biography of Wolfgang Pauli

by Charles P. Enz

Oxford University Press: 2002. 581 pp.

£35, \$60

Freeman Dyson

Wolfgang Pauli was one of the central figures in the history of twentieth-century science. His discovery of the exclusion principle in 1924 provided the essential foundation for the development of modern chemistry and molecular biology. This was the discovery for which he received a belated Nobel Prize in 1945. He also made three discoveries of comparable importance for the development of physics: the Pauli spin-matrices, the theory of the neutrino, and the spin-statistics theorem, which states that particles obey the exclusion principle if, and only if, they have 'half-integer' spin.

In addition, Pauli also wrote review articles summarizing various fields of physics in a masterly fashion, beginning with his legendary encyclopedia article on relativity, published in 1921. When Einstein reviewed this article, he wrote: "No one studying this mature, grandly conceived work would believe that the author is a man of twenty-one." Pauli was determined to understand and criticize all of the new ideas that were emerging in the rapidly growing physics of his time. His criticisms were accurate and often merciless, and earned him the title 'the conscience of physics'. When Pauli condemned a piece of work as trivial or sloppy, his verdict was final. His friend Paul Ehrenfest called him 'God's whip'.

Charles Enz has chosen an apt title for this biography: he was determined to understand and explain every paper that Pauli published. He gives us not only the ideas, but also the equations, sparing neither space nor time to get the details right. He also summarizes numerous papers by other authors, providing the context for Pauli's contributions. About one-sixth of the book consists of end-notes containing references to the relevant literature. Enz has faithfully reconstructed Pauli's scientific output, with every brick of the edifice in its proper place.

Interspersed among these technical expositions are narrative passages that describe Pauli's personal life and activities outside science. For most readers, the narrative passages will be the most interesting part of the biography, but they add up to much less than half of it. Quite apart from being a great scientist, Pauli was a great character and an acute observer of the human scene. He lived through tragic times and had friends in many

countries. He was an indefatigable writer of letters, and always expressed himself with wit and style. Several volumes of his correspondence have been published, edited by Karl von Meyenn, and there is ample material for a more extended narrative of Pauli's life. I hope that Enz might one day write a second biography, written for readers who are not interested in equations, giving a broader view of Pauli and his friends, and the history through which they lived.

Enz was the last of Pauli's gifted young assistants. In the European academic system, each professor had an assistant who acted as a personal factotum. The assistant corrected proofs and examination papers, gave the professor's regular lectures when the professor was away, and was expected to have enough time left over to do independent research. Pauli loved to talk, and one of the duties of his assistant was to listen. Pauli mellowed at the end of his life, and his routine became more relaxed. He liked to go for leisurely walks after lunch, stopping for ice-cream and extended conversations on the way. Enz became a close friend of Pauli and his wife Franca, and helped her to deal with Pauli's papers after his death. He probably knew Pauli, both on a professional and a personal level, better than anyone else still living.

The most unexpected side of Pauli's life was his involvement with the psychiatrist Carl Jung. They got to know one another in 1932 when Pauli was seriously depressed after his first marriage had broken up two years earlier. Jung treated Pauli by analysing his dreams. Jung studied 1,300 of Pauli's dreams and used them as raw material for his speculative theories of archetypes and the collective unconscious. For Jung, Pauli was the ideal patient, providing an endless supply of episodes and images with which to explore the unconscious. For Pauli, Jung was the ideal physician, turning his personal problems into an intellectual discussion that they could both enjoy. The exercise of writing down the dreams and discussing their interpretation with a sympathetic listener soon relieved Pauli's depression.

Within two years, Pauli was married to Franca and the treatment came to an end. Pauli and Jung remained close friends, and Pauli continued to write down his dreams, occasionally sending them to Jung. In 1952 Pauli published a historical paper, "The Influence of Archetypal Ideas on the Scientific Theories of Kepler", which was heavily influenced by Jung. And two of Pauli's more striking dreams were described by Jung in his 1944 book *Psychology and Alchemy*. Jung's theories are today widely regarded as



Great minds: Wolfgang Pauli (right) with Albert Einstein in 1926

mystical mumbo-jumbo, but Pauli took them seriously and gave to them the same critical scrutiny that he gave to physics.

In Pauli's view, the secrets of nature were to be found in human understanding as much as in science, and Jung's theories were essential keys to human understanding. In this, Pauli was following the example of Isaac Newton, who devoted much of his time and effort to alchemy and theology, as well as to the discovery of the basic laws of physics.

Another bizarre aspect of Pauli's life was his relationship with Switzerland. A native of Austria, he settled in Zürich in 1928 and lived there until 1940 when he moved to Princeton, New Jersey. He felt at home in Switzerland and left only because the authorities refused to give him Swiss citizenship. In 1945 the Institute for Advanced Study in Princeton offered him a permanent professorship, which he declined because he felt lonely and out of place in the United States. In a letter to a Swiss friend, Pauli wrote: "As in Austria during the First World War, in this year in the USA I suddenly had the feeling that I was placed in a criminal atmosphere ... at the time when those A-bombs were dropped." In 1946 Pauli became a naturalized US citizen in order to have a passport with which to travel back to Switzerland. In 1949, after many applications and long arguments, he

was finally granted Swiss citizenship.

This biography is hard to read but is worth the trouble. One of Pauli's friends, the mathematician Hermann Weyl, wrote in the preface to his book *The Classical Groups*: "The gods have imposed upon my writing the yoke of a foreign tongue that was not sung at my cradle... Nobody is more aware than myself of the attendant loss in vigour, ease and lucidity of expression." The many quotations from Pauli in this book suffer from the same loss. They are translated from Pauli's racy and idiomatic German into stilted and unnatural English. Yet despite the difficulties of translation and the prevalence of highly technical mathematics, Enz has given us an authentic portrait of one of the great spirits of our time. ■

Freeman Dyson is at the Institute for Advanced Study, Princeton, New Jersey 08540, USA.

.....

Through the eyes of God's naturalist

Glimpses of the Wonderful: The Life of Philip Henry Gosse

by Ann Thwaite

Faber & Faber: 2002. 407 pp. £25

Rebecca Stott

In 1857 Charles Darwin wrote to the distinguished zoologist Philip Henry Gosse to request information about "crustacean battles", asking: "Can you tell me, you who have so watched all sea-nature..." Darwin's address to his fellow naturalist is a testimony to Gosse's standing in the history of natural history. Gosse's eyes, pressed to the lens of a microscope, peering into rock pools or through the glass of his window-sill aquarium, had indeed seen more sea nature than perhaps those of any other man or woman in England by 1857. In his lifetime Gosse published a staggering 39 books, all but 10 of which were on natural history, and many were best-sellers. On his death the Royal Society claimed that "no man has ever done so much to popularise the study of natural history in England".

The publication of his son Edmund's poignant autobiography *Father and Son: A Study of Two Temperaments* (Heinemann, 1907) transformed Philip Gosse from a distinguished Victorian zoologist into a legend. The Gosse legend, as shaped by father and son, takes two parts. First, Philip is cast as a repressive evangelical who relentlessly and cruelly scrutinized his son's soul as if under a microscope. Second, he is the tragic victim of his own evangelical beliefs, attempting in his book *Omphalos* (Van Voorst, 1857) to reconcile science and religion by claiming that God planted fossils in the rocks because he wanted to create a world with the appearance of a prehistory.



PHILLIP HENRY GOSSE

Living colour: one of Gosse's plates from *The British Sea Anemones and Corals* (Van Voorst, 1860).

Ann Thwaite's remarkable biography takes Gosse out of melodrama and into realism. In Thwaite's hands he is complex and in conflict: at once scientifically curious yet doctrinally absolute, charming and vehement, driven and tormented by the apparent conflict between his belief in the fixity of species and the assimilation of contrary evidence in zoology. It is a fascinating portrait, delicately drawn and moving.

The book's title, *Glimpses of the Wonderful*, is a reference to the title of one of Gosse's first nature books, and underscores the fact that this is a book as much about ways of seeing and reading nature in the nineteenth century as it is a biography of Gosse himself. His way of seeing was the microscopic vision of the describer and systematist; he lacked the telescopic and speculative vision of the philosophical naturalist concerned with the origins and transmutations of species. Darwin complained that "in Gosse's books there is not enough reasoning for my taste", but Gosse wrote that "what I delight in is the minute details of habits, the biography of animals". Reasoning was best left to others; Gosse had in its stead rare gifts of observation and description.

For over 60 years Gosse studied, catalogued, sketched, painted and put into words the tentacled, often erotic, anatomically bizarre life of the seabed or pond, as well as writing on butterflies, birds and orchids. He denounced the natural-history practice of many of his peers as a 'necrology', the science of dead things, stuffed and pinioned. Instead he sought to bring his biographical subjects, his 'glimpses of the wonderful', poetically to life in a series of word pictures. The books are vividly described (rotifers, for instance, are "creatures that swim with their hair, that have ruby eyes blazing deep in their necks"), and his hand-painted illustrations are now collectors' items.

Gosse was rarely still except in prayer. A fellow naturalist wrote in the 1870s: "I can picture to myself Gosse plunging into a pool in full sacerdotal black, after a sea anemone." He farmed in Canada, stalked birds in Jamaica, worked as a schoolteacher in London, translated, wrote, studied, painted and preached. Thwaite's Gosse is a haunted man, driven both by curiosity and a keen sense of having little time left. As a member of the Brethren, a group of Christians dedicated to keeping the principles of the early Christian church, Gosse waited for the return of Christ in Rapture, watching for signs of the end, interpreting prophesy. If the Rapture were both imminent and longed-for, a book might be left unfinished or a sea anemone unclassified when the time came. As a result he was prolific until his final years. Edmund Gosse, visiting his father towards the end of his life, wrote: "I could see for the first time that his vehement eager life would not last for ever."

Histories of nineteenth-century nature observation confirm that the act of seeing and describing is shaped by ideological convictions. Thwaite's book gives us a greater insight into the complexity of Victorian ways of seeing nature by showing that Gosse was far from blinded by his conviction that species were fixed and God-created; he could see in the lens of his microscope the increasingly magnified evidence to the contrary. It is a testimony to his sincerity that he tried to square the circle in *Omphalos*. His tragedy was that the circle could no longer be squared. ■

Rebecca Stott is reader in Victorian Literature and History at Anglia Polytechnic University, East Road, Cambridge CB1 1PT, UK, and is an affiliated scholar in the Department of the History and Philosophy of Science, University of Cambridge, Free School Lane, Cambridge. She is the author of the forthcoming book *Darwin and the Barnacle* (Faber & Faber).

So near, but yet so far

Human Evolution Through Developmental Change

by Nancy Minugh-Purvis & Kenneth J. McNamara

Johns Hopkins University Press: 2002. 536 pp. \$59.95, £40

Bernard Wood

"Are humans really our closest ancestors?" If there were a newspaper for chimpanzees, this might well be one of its headlines. But there are no chimp news barons because the phenotypes and behaviours of chimps and modern humans are critically different. Yet many lines of evidence suggest that chimps and humans evolved from a common ancestor that lived between 5 million and over 10 million years ago. The challenge faced by those engaged in human evolution research is to reconstruct that common ancestor and then flesh out the details of its transformation into modern humans.

The common ancestor of chimps and humans was, of course, neither a chimp nor a modern human. But there are good reasons to think that many aspects of living chimps, including the timing and the sequence of events in their growth and development, such as weaning infants or changes in dentition, are reasonable proxies for the primitive condition in the common ancestor of chimps and humans.

Evolutionary change is brought about by altering growth and development. These

changes can be restricted to a single structure, such as the hand, or they can be more or less global in their scope. These more general changes to the relative timing and the rate of development are subsumed into the term 'heterochrony'. Speeding up or slowing down the growth of the whole, or a substantial part, of an animal can change both the size and the shape of the evolved descendant.

Size, shape and timing are relatively crude ways of comparing ancestors and descendants, but they have the advantage of being relatively easy to measure, and we can translate them into changes in the onset, duration and offset time of cellular activity. Until molecular developmental biologists can point to the specific genetic mechanisms that determine the way in which an individual's genome is translated into its phenotype, we may have to be satisfied with comparing the results of cellular activity, such as dental microstructure.

In this volume the editors address three main topics. The first is a broad-brush review of the links between development and evolution. The second is advertised as the evolution of life history within the higher primates, but in effect it concentrates on the developmental changes that underline phenotypic differences between living primates, with only a few intrepid authors tackling the thorny problem of reconstructing the evolutionary history of these differences. The third and final section concentrates on a few specific examples of how understanding the developmental sequence can inform hypotheses about the developmental basis of morphological differences between fossil hominin species.

Several themes emerge from the diverse contributions. The first is the confirmation that modern humans did not evolve from a more chimp-like precursor, because of some relatively simple global retention of juvenile characteristics. Distinctive sets of transformations seem to underlie each major change in morphology or life history. The second theme is the emphasis given in the volume to the evolution of the distinctive life history of modern humans. The pace and tempo of development is very different in modern humans and chimps. Like most primates, chimps have a general retardation of development (an extended infancy), but modern humans have taken that trend to apparently unique levels of ontogenetic sloth. Even when the sequence of developmental events is the same, modern humans track through those stages much more slowly than chimps. Indeed, we have so effectively brought forward the time of weaning from the six or seven years we would predict from our body mass to about three years that we have had to invent a new term — childhood — to describe the stage in our life history between weaning and the end of brain growth and the eruption of the first molar teeth.

The third theme is the comprehension of just how difficult it will be to trace the evolutionary history of our unique phenotype and life history. Several studies of fossil evidence demonstrate that the phenotypic differences between modern humans and Neanderthals are established early in development. But it will be a long time before we have comparably sized samples for the taxa that are found earlier in the human fossil record. Only then will we have enough evidence to afford researchers equivalent opportunities to trace the ontogeny of what we hypothesize are taxonomically significant morphological differences between early hominin species.

Working out when, and in what context, the various distinctive aspects of modern human morphology and life history evolved is a formidable challenge. But unless that challenge is accepted and faced, we will remain ignorant about the evolutionary history and adaptive context of many of the most important components of our humanity. ■

Bernard Wood is in the Department of Anthropology, George Washington University, Washington DC 20052, USA.

More on human evolution The Speciation of Modern *Homo sapiens*

edited by T. J. Crow

Oxford University Press, £29.50

The Neanderthal's Necklace: In Search of the First Thinkers

by Juan Luis Arsuaga

Four Walls Eight Windows, \$25.95



The dance of chance

Does God play dice? Is reality slave to chance? British artist Tim O'Reilly responded to his confrontation with quantum theory during a recent visit to CERN, the European laboratory for particle physics, with a video installation of

two dice orbiting each other in an endless loop, postponing indefinitely the chance moment of their fall. The video is part of the "Signatures of the Invisible" exhibition at the Gulbenkian Museum in Lisbon, which runs until 19 January.

Aspirational thinking

Henry Gee

We've all seen the commercials. A line of figures walking from left to right, first a shambling ape on all fours; the second, semi-erect with a vague glimmer of intelligence, and perhaps holding a hand-axe; further along, a tall, proud man, carrying a spear and wearing furs; and finally, a user of the latest car or washing machine. The caption will speak of advancement and progression, something like 'Evolution — the Next Step'.

These advertisements reflect the popular view of evolution as a progressive force that drives an inexorable improvement. But in reality, evolution is based largely on natural selection, a handy term for the interaction between the environment, mutation and superabundance. Natural selection has neither memory nor foresight; it works only in the here and now. It is not a force, an entity separate from the materials on which it acts. Still less can it be personified.

Although one might like to blame the advertising industry for this misrepresentation, a successful copywriter will only hold a mirror up to the zeitgeist, and popular wisdom sees evolution as progressive and directed. So why, almost a century and a half after Darwin, do we still so readily accept this view of evolution as progressive?

I blame nature philosophy, a remarkable movement that flowered in Germany in the eighteenth century, and whose adherents were both acutely scientific and breathlessly romantic at the same time. In nature philosophy, all organic forms are manifestations of a cosmic compulsion towards perfection, with the human form as its ultimate destiny. As Lorenz Oken (1779–1851) put it: "What is the animal kingdom other than an anatomized man, the macrocosm of the microcosm?" No copywriter could have put it better: those evolving-human commercials are just

nature philosophy brought up to date.

The most famous nature philosopher was Johann Wolfgang von Goethe (1749–1832). Although best known as a poet and dramatist, his scientific achievements were considerable and he had a profound influence on nineteenth-century biologists. Thomas Huxley's article "Nature: Aphorisms by Goethe" appeared in *Nature's* first issue in 1869. And yet in Goethe we see the apotheosis of nature philosophy as a romantic reaction to what we would see as scientific detachment, seeking to place man, once again, at the centre of all things, and to promote the subjective and the aesthetic in scientific observation.

Goethe's wide range of interests led to his being labelled as an amateur by contemporaries who had a narrowly scientific focus. In Goethe's own words, his critics "forgot that science arose from poetry, and did not see that when times change the two can meet again on a higher level as friends". Although nature philosophy is long dead, such sentiments still find ready acceptance among alternative or 'holistic' philosophies. Anthroposophy — the world view of twentieth-century philosopher Rudolf Steiner — draws heavily on Goethe, and a germ of nature philosophy survives, if buried, in every anti-scientific, anti-establishment eco-warrior. Why has nature philosophy reinforced the idea of progressive evolution, given that it stems from a profoundly idealistic, pre-evolutionary view of life?

The unlikely fusion between evolution and nature philosophy was brokered by the zoologist Ernst Haeckel (1834–1919), who revered Goethe as a founding father of comparative anatomy. Haeckel was a product of the nineteenth-century school of German embryology that had been founded by nature philosophers, and it is largely thanks to him that we can now make, almost

Progressive evolution

Those evolving-human commercials are just nature philosophy brought up to date.

without thinking, the connection between the embryogeny of individual organisms and the grand sweep of the progress of life on Earth. This link might not have been made as explicitly — or even at all — without nature philosophy. Indeed, one of Haeckel's students was Wilhelm Roux (1850–1924), a founder of modern developmental biology.

Haeckel visualized Darwin's natural selection as the engine driving a kind of evolution that ran on progressive, improving lines — a kind of animated nature philosophy. This line of thought produced, in other hands, the popular-culture model of evolution in which humans are the goal and the final statement.

Of course, natural selection is not a force, like gravity. It is directionless with respect to history; if there is direction in evolution (perhaps biased by developmental constraint), it is not propelled by any inherent drive for improvement. So why have modern scientists and so-called science popularizers failed to establish this truth in popular culture?

An obvious reason is that the progressive view resonates far more strongly with our own vanity and inclinations than with the more abstract and austere concept of evolution by mindless selection. In less enlightened times, progressive evolution was used to justify racism and Nazism. We like to think that we have risen above such things, but the copywriters know better — when we see the canonical parade of evolving humans, we identify with and aspire to be the one at the top.

Perhaps a more subtle explanation is that the progressive view is lodged in the minds of senior advocates of darwinian evolution by natural selection. As the late J. Z. Young wrote: "We shall expect to find in mammals even more devices for correcting the possible effects of external change than are found in other groups ... culminating in man with his astonishing perception of the 'World' around him and his powers of altering the whole fabric of the surface of large parts of the earth to suit his needs." Perhaps there is a nature philosopher in us all. ■

Henry Gee is a Senior Editor of *Nature*.

FURTHER READING

Young, J. Z. *The Life Of Vertebrates* 3rd edn (Oxford Univ. Press, New York, 1981).
Huxley, T. H. *Nature* 1, 9–11 (1869).



Hair today, gone tomorrow: popular culture still labours under the misapprehension that evolution is an inexorable climb culminating in humans, the last word in sophistication and sartorial savoir-faire.

On the wings of inhibition

Jeffrey L. Wrana

Growth factors regulate cell behaviour, and are kept in check by inhibitors. The structure of a complex of two such proteins shows that they form back-to-back butterflies, with the inhibitor's wings stretching to embrace its partner.

Multicellular organisms use a balance of secreted growth-factor proteins and their inhibitors to coordinate cell behaviour, thereby regulating the development of embryos and maintaining the status quo in adult tissues. On page 636 of this issue, Groppe and colleagues¹ describe the structure of one such growth factor, bone morphogenetic protein 7 (BMP-7), in a complex with its inhibitor, the Noggin protein. This is the first glimpse of a member of the BMP family bound to its antagonist, and it shows that Noggin has a strikingly similar structure to that of the growth factor it inhibits. This surprising observation implies that the two proteins may have arisen from a common ancestral gene.

If single-celled embryos are to accurately and reproducibly grow and develop into multicellular adult animals, the cells that are generated during this process must coordinate their proliferation, movement, specialization and even death. This teamwork requires that cells talk to each other, and one way in which they do this is to emit and respond to both short-range and long-range signals that direct cell behaviour. Some of the major signals are the secreted growth-factor proteins, of which one class is the transforming growth factor- β (TGF- β) superfamily. The BMPs form a subgroup of this super-

family that was first defined on the basis of an unusual biological activity: when implanted in a suitable matrix, BMPs can induce bone formation in muscle². In addition to being key regulators of bone and joint formation, however, these proteins regulate a vast array of other processes, both during development and in adults³.

Each BMP is composed of two identical proteins (monomers), which are linked into a dimer by means of disulphide bonds between cysteine amino acids. The BMPs work by binding to a receptor complex that is found on the surface of almost all normal cells, and is composed of so-called type I and type II receptors. Structural studies^{1,4,5} have indicated that each monomer of a BMP (and of TGF- β itself) can connect to both the type I and the type II receptor (Fig. 1a). This arrangement probably supports the assembly of a tetrameric receptor complex, comprising two type I receptors and two type II receptors (Fig. 1b), which is essential for transducing the signal into the cell.

Secreted growth factors such as BMPs are powerful regulators of cell function. But they are also independent agents that do not discriminate among the cells with which they come into contact. Consequently, an elaborate network of secreted inhibitors is used to ensure that growth-factor activity is

restricted to the right place, the right time and the right cell types. In this way, a developing organism keeps its growth factors on a tight leash. Several types of BMP inhibitors are found in vertebrates (see references in ref. 1). One of these is Noggin. As the name might suggest, this protein affects head development: it was discovered in screening tests for genes that could 'rescue' head formation in embryos of the African clawed frog (*Xenopus laevis*) that had been pretreated to prevent normal head development⁶. Only one Noggin gene has been found so far in mammalian genomes, but it is essential in humans: mutations in just one of the two copies of the gene lead to skeletal dysplasias (developmental disorders of the bone and cartilage) that are characterized by joint fusions⁷. So, understanding how antagonists such as Noggin function at the molecular level is an important topic in both developmental biology and human disease.

Groppe *et al.*¹ determined the structure of Noggin bound to BMP-7 by using X-ray crystallography. Analysis of the structure revealed remarkable similarities between the two proteins. Like BMP-7, Noggin forms a dimer with a two-fold axis of symmetry. The dimer is shaped like a butterfly, with structural features known as antiparallel β -sheets in each monomer extending wings from a core

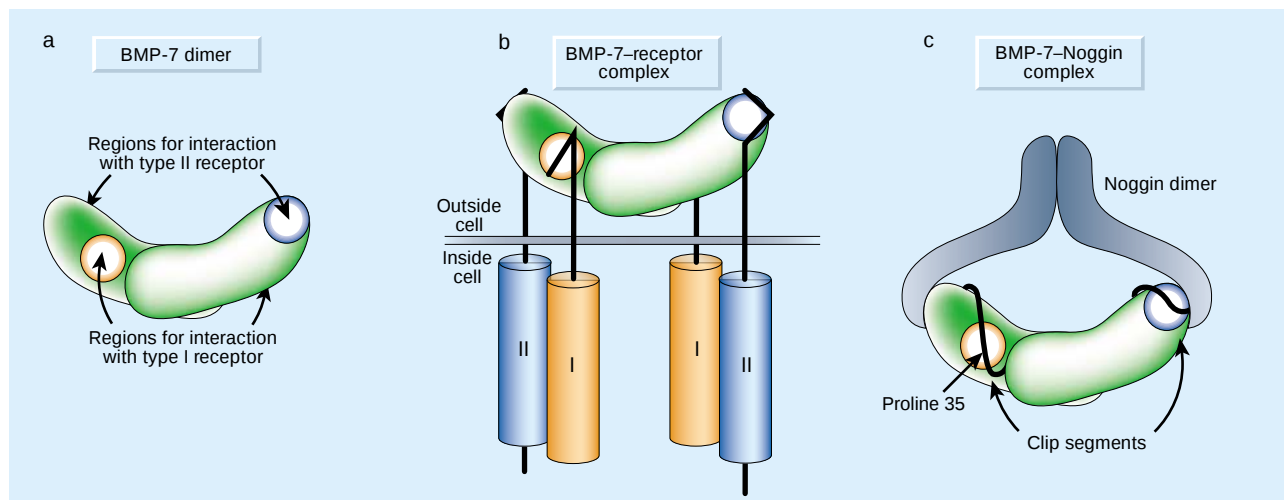


Figure 1 Antagonizing growth factors. a, Representation of the BMP-7 (bone morphogenetic protein 7) dimer, showing surfaces that interact with type I (orange) and type II (blue) receptors. b, How BMP-7 might bind to a tetrameric complex of type I and type II receptors, which feed the signal from BMPs into cells and thereby influence cell behaviour.

c, Representation of the structure of BMP-7 with its antagonist Noggin, solved by Groppe *et al.*¹. The diagram shows the amino-terminal clip segment of Noggin that occludes the receptor-interaction surfaces of BMP-7, and the position of proline 35, which inserts into the hydrophobic pocket of BMP-7 that makes contact with a phenylalanine in the type I receptor.

body. The core body mediates dimerization and forms a cystine knot; this is a characteristic structure of disulphide-bonded cysteines that is also found in BMP-7 and numerous other secreted proteins⁸. The knot takes the general shape of a ring of intramolecular disulphide bonds, through which passes a single intermolecular disulphide linkage that stabilizes the dimeric structure.

The BMP-7 dimer has wings that are structurally similar to but more compact than those of Noggin. In the complex¹, the Noggin and BMP-7 butterflies are arranged back-to-back with the wings touching (Fig. 1c) and Noggin's wings extending out to embrace those of BMP-7. This overall structure provides an essential spatial arrangement that allows an amino-terminal extension of each Noggin monomer to snake around BMP-7 and form a 'clip' that occludes the surfaces of the growth factor that make contact with its receptor. Indeed, one of the amino acids in the amino-terminal half of the Noggin clip (proline 35) extends into a hydrophobic pocket on BMP-7 that normally⁴ makes a key contact with a phenylalanine residue on the type I receptor. At the other end of the clip, hydrophobic amino acids cooperate with others to mask the hydrophobic patch on BMP-7 that makes contact with the type II receptor. These extensive contacts and the occlusion of both receptor-binding sites on BMP-7 probably explain both the high affinity of Noggin for BMP-7 and its potent antagonistic activity.

So just how important are these contacts to Noggin's antagonism of BMP-7 and, ultimately, to its physiological functions? To answer this question, Groppe *et al.* produced versions of Noggin that had different mutations in each of the interaction surfaces, and used them in tests of limb-bud development in the chick. In normal limb development, BMPs are essential for formation of the cartilage that presages where bones will form⁹. They are also needed to sculpt the fingers, by inducing cell death between the digits. So, treating developing limb buds with Noggin leads to a loss of cartilage formation and a block in cell death between digits. But Groppe *et al.* found that mutations in Noggin's second interaction surface strongly reduced its ability to bind BMP-7 and to antagonize BMPs *in vivo*. In contrast, mutation of the first interaction surface had a less dramatic effect on binding to BMP-7, and weak effects on limb development. Nevertheless, this latter mutation does cause joint fusions in a number of human skeletal dysplasias, so the affected surface is biologically important¹.

This Noggin-BMP-7 structure¹ provides the first glimpse of how a secreted antagonist blocks BMP function. It remains to be seen whether BMP antagonists from another, larger protein family — the DAN family — have similar overall structures. But at least

one such antagonist, Cerberus, can inhibit BMPs as well as members of another, structurally unrelated family of secreted factors, the Wnt proteins¹⁰. It will be interesting to see how the DANs can block multiple pathways. Furthermore, Noggin and DAN genes are rarely found in invertebrate genomes, but are conserved in diverse vertebrate species. The finding that Noggin and BMP-7 share considerable structural similarity, despite having opposite biological functions, suggests that the demands of patterning vertebrate organisms may have driven the evolution of a BMP antagonist from an ancient BMP-like gene.

Jeffrey L. Wrana is in the Department of Medical

Genetics and Microbiology, University of Toronto, and the Program in Molecular Biology and Cancer, Samuel Lunenfeld Research Institute, Mount Sinai Hospital, 600 University Avenue, Toronto M5G 1X5, Canada.

e-mail: wrana@mshri.on.ca

1. Groppe, J. *et al.* *Nature* **420**, 636–642 (2002).
2. Urist, M. R. *Science* **150**, 893–899 (1965).
3. Hogan, B. L. M. *Genes Dev.* **10**, 1580–1594 (1996).
4. Kirsch, U., Sebald, W. & Dreyer, M. K. *Nature Struct. Biol.* **7**, 492–496 (2000).
5. Hart, P. J. *et al.* *Nature Struct. Biol.* **9**, 203–208 (2002).
6. Smith, W. C. & Harland, R. M. *Cell* **70**, 829–840 (1992).
7. Gong, Y. *et al.* *Nature Genet.* **21**, 302–304 (1999).
8. Vitt, U. A., Hsu, S. Y. & Hsueh, A. J. *Mol. Endocrinol.* **15**, 681–694 (2001).
9. Capdevila, J. & Izpisua Belmonte, J. C. *Annu. Rev. Cell Dev. Biol.* **17**, 87–132 (2001).
10. Piccolo, S. *et al.* *Nature* **397**, 707–709 (1999).

Nuclear physics

A triple point in nuclei

David Warner

Triple points describe states of matter in which three phases exist at the same time — such as solid, liquid and gas. The same phenomenon has now been found to occur between three different shapes of atomic nuclei.

The most familiar example of a triple point is in water: on a graph of pressure and temperature, the three lines separating the vapour–liquid, liquid–solid and solid–vapour phases all cross at a temperature of 273.15 K, that value being set to define the Kelvin scale of temperature. Phase transitions such as these are normally associated with temperature, but on the quantum scale they can occur at zero temperature through other mechanisms. In the lowest energy (ground) states of atomic nuclei, different phases exist which correspond to different geometrical shapes. According to Jan Jolie and colleagues¹, writing in *Physical Review Letters*, these phases come together at a triple point, validating a prediction made by the Russian physicist Lev Landau in a classic

paper on the theory of second-order phase transitions². Writing in 1937, Landau could not have expected his theory to apply in the nuclear domain.

When ice melts, it does so at a particular temperature and with a sudden change in its state, as the crystal structure is destroyed by the thermal motion of the water molecules. This is an example of a first-order phase transition: at the transition temperature, the two phases — solid and liquid — coexist and latent heat is required to effect the transformation from one to the other. Landau's theory deals with second-order phase transitions in which the state of the system changes in a continuous way with no coexistence of phases. Instead, at the transition point, the two phases become

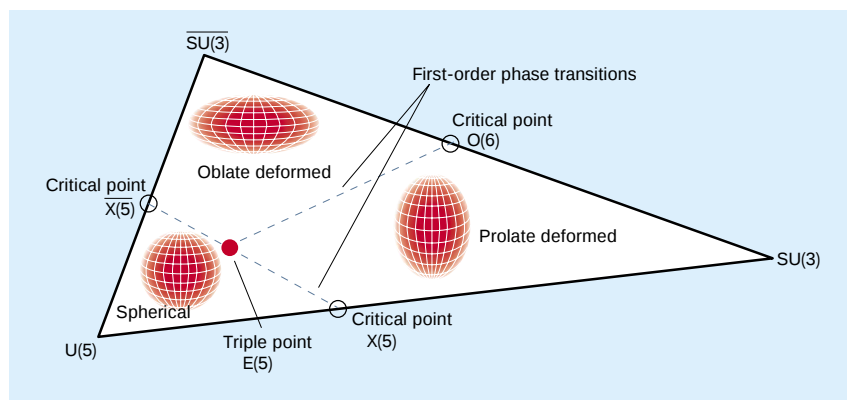


Figure 1 The extended Casten triangle. Each apex denotes a mathematical symmetry corresponding to one of the three shapes shown. Transition points and their associated critical symmetries are indicated, as are first-order phase transitions. Jolie *et al.*¹ propose that there is a nuclear triple point that marks the second-order transition between a spherical nuclear shape and a prolate or oblate deformed one. Existing data for the barium nucleus support this hypothesis.

indistinguishable. A familiar example is the magnetic transition in iron, in which the magnetization associated with the ferromagnetic state vanishes at the Curie temperature and the system becomes paramagnetic.

The phase transitions in these classical systems arise from a competition between order and thermal fluctuations. In quantum systems, however, phase transitions can take place at zero temperature and the change in order is invoked by some other parameter. In its ground state, an atomic nucleus has a stable geometrical shape that results from the interaction between its constituent neutrons and protons (known collectively as nucleons). As the number of nucleons changes from nucleus to nucleus, shape phase transitions occur, in which the geometrical configuration changes, for example, from spherical to 'prolate quadrupole deformed', or cigar-shaped. (Strictly speaking, phase transitions in finite systems can only be defined in the classical limit in which the number of constituents tends to infinity. In practical terms, this means that the discontinuity associated with the transition is smoothed out in the finite system.) Quantum phase transitions involving changes in geometrical configurations occur in other finite many-body systems as well, such as molecules.

Interest in nuclear-shape phase transitions has been galvanized over the past two years by a new interpretation of the behaviour of the nuclei ^{152}Sm (samarium) and ^{134}Ba (barium). In nuclear physics, quantum phase transitions can be studied most easily using algebraic techniques that associate a specific mathematical symmetry with the different nuclear shapes. In particular, in the framework of the 'interacting boson model'³, three such shapes exist corresponding to spherical symmetry, a deformed shape with axial symmetry and a deformed shape without axial symmetry (described mathematically by the groups $U(5)$, $SU(3)$ and $O(6)$, respectively). The transition between the first two of these is represented by a line of first-order transitions on the phase diagram, and the transition between $U(5)$ and $O(6)$ is a second-order point.

New experimental data^{4,5} and a fresh examination⁶ of both data and theory for ^{152}Sm and ^{134}Ba have demonstrated the relevance of the concept of quantum shape phase transitions in the nuclear domain. Moreover, Francesco Iachello has developed 'critical' symmetries that provide an analytical way to describe the structure of nuclei at the critical points of the second-order transition⁷ and in the coexistence region of the first-order transition⁸ ($E(5)$ and $X(5)$, respectively) — predictions that are again matched by the behaviour of ^{152}Sm and ^{134}Ba .

The picture was not complete, however, without the new work published by Jolie *et al.*¹. Until now, the nuclear phase diagram, known as the Casten triangle (Fig. 1), had

shown a line of first-order phase coexistence between the $U(5)$ and $SU(3)$ symmetries that culminated in a second-order point defining the $U(5)$ – $O(6)$ transition. But where was the transition in the third leg of the triangle, between $SU(3)$ and $O(6)$? In earlier work, Jolie *et al.*⁹ showed that part of the triangle was missing: it should be extended to incorporate $SU(3)$ symmetry. This corresponds to axially symmetric, oblate deformation, instead of the prolate shapes represented by $SU(3)$ — a flattening rather than a stretching, essentially a change in sign of the relevant shape variable. The previously established $O(6)$ symmetry then also plays the role of a critical symmetry, describing the first-order phase transition from a prolate to oblate shape.

Armed with this extended version of the Casten triangle, Jolie *et al.*¹ have now completed the story by analysing the nuclear-shape phase diagram in terms of the Landau theory of continuous (second-order) phase transitions. The authors show that, in the nuclear case, a second-order transition can only occur as an isolated point that coincides with a junction of two lines of first-order transitions (Fig. 1). This is the triple point of nuclear deformation. In agreement with Landau's theory, the phase transitions occur between symmetries of higher and lower order (spherical and deformed) and between symmetries characterized by opposite signs

of the order parameter (corresponding to prolate and oblate deformation). At the triple point, all three phases exist and, looking back to earlier experimental studies, the ^{134}Ba nucleus has provided the first experimental proof that this is so.

The study by Jolie *et al.*¹ shows, in a straightforward and elegant way, how the classical theory of phase transitions can be applied to describe the shapes of atomic nuclei in their ground states. The theory describes the nature of the transition from one shape to another as a function of the appropriate parameters — a unified picture that offers a new perspective on the changing shape of nuclei. It also emphasizes how basic theoretical concepts can span a multitude of physical systems. ■

David Warner is in the Surface and Nuclear Division, Daresbury Laboratory, Daresbury, Warrington, Cheshire WA4 4AD, UK.
e-mail: d.warner@dl.ac.uk

1. Jolie, J. *et al.* *Phys. Rev. Lett.* **89**, 182502 (2002).
2. Landau, L. *Collected Papers of L. D. Landau* (ed. ter Haar, D.) 193–216 (Pergamon, Oxford, 1965).
3. Iachello, F. & Arima, A. *The Interacting Boson Model* (Cambridge Univ. Press, 1987).
4. Casten, R. F. & Zamfir, N. V. *Phys. Rev. Lett.* **87**, 052503 (2001).
5. Casten, R. F. & Zamfir, N. V. *Phys. Rev. Lett.* **85**, 3584–3586 (2000).
6. Iachello, F., Zamfir, N. V. & Casten, R. F. *Phys. Rev. Lett.* **81**, 1191–1194 (1998).
7. Iachello, F. *Phys. Rev. Lett.* **85**, 3580–3583 (2000).
8. Iachello, F. *Phys. Rev. Lett.* **87**, 052502 (2001).
9. Jolie, J. *et al.* *Phys. Rev. Lett.* **87**, 162501 (2001).

Aerodynamics

Red admiral agility

Rafał Żbikowski

Our understanding of insect flight is hampered by the difficulty of obtaining data when the insects are flying freely. But such experiments can be carried out and show butterflies to be masters of flight control.

Over the past ten years there has been much progress in understanding insect flight^{1–3}. The news keeps coming, and the latest instalment appears on page 660 of this issue⁴, where Srygley and Thomas describe aerodynamic observations made on free-flying red admiral butterflies. By injecting wisps of smoke into a wind tunnel, through which the butterflies flew towards an artificial flower, the authors were able to record airflow features around the wings using high-speed cameras. From the ensuing analysis, they conclude that the butterflies possess an impressive repertoire of aerodynamic mechanisms, which they employ in different circumstances and with great virtuosity.

We evidently still have much to learn about insect flight, which has been refined over an evolutionary history of 300 million years or more⁵. There are plenty of reasons for such studies. This is a subject of fundamental importance in biomechanics⁶, and

more generally in biology, because of the abundance and ubiquity of insects, and their importance in many ecosystems⁷.

Insects are also masters of manoeuvrability at low speeds, in hovering and flying backwards and sideways, skills that are of great interest to engineers in two respects. One is the lessons to be learned in designing flapping-wing 'micro air vehicles' (defined as being no more than 15 centimetres in length, width or height). Another is the prospect of reverse-engineering insect flight control, by finding out how insects steer in the air. Man-made flying vehicles are controlled by software commands, but the software design requires many man-years of work and powerful computer chips for its implementation. By contrast, in flies for instance, flight control probably originates from a central complex in the fly brain consisting of about 3,000 neurons⁸. This gives the insect less computational power than a toaster, yet insects are more agile than aircraft equipped



100 YEARS AGO

The services which photography has rendered to science are now well recognised, and its value for purposes both of observation and record is well known and admitted. It is probably not so well known that methods now exist by which not only the form, but the colour, of natural objects can be represented with approximate fidelity. We are fortunate in being able to illustrate this fact by a plate giving some excellent reproductions of birds' eggs, produced under the superintendence of Mr. H. E. Dresser, entirely by photographic methods, and without the intervention of an artist. There is no need to dwell on the value of such work. For many scientific purposes it is as important to record colour as shape, and if this can be done in a trustworthy manner, a new and useful power is placed at the disposal of the teacher of science and of the writer of scientific books. The difficulty about the three-colour process of photography is that it is extremely difficult to make certain that the colours are reproduced with sufficient accuracy for scientific work. Accuracy enough for pictorial purposes is easily attained, but absolute truth to nature is quite another thing.

From *Nature* 11 December 1902.

50 YEARS AGO

Born in Paris a century ago, on December 15, 1852, Antoine Henri Becquerel came of a family long distinguished in the world of physical science. Educated at the École Polytechnique, Paris, his early studies were concerned with the magnetic rotation of the polarization plane of light... and the absorption of light by crystals. In 1892 he was appointed professor of physics at the Musée d'Histoire Naturelle — a chair held before him by his father, Alexandre Edmond Becquerel (1878), and by his grandfather, Antoine César Becquerel (1837). Three years later he became professor of physics at the École Polytechnique. His greatest achievement was the result of a lucky accident, but at the same time it was the culmination of a long series of carefully planned experiments. In 1896 he found that photographic plates, protected against ordinary actinic radiations, were fogged by emanations from uranium ores. His paper entitled "Sur les radiations émises par phosphorescence"... ushered in the new era of radioactivity. Their discoverer in 1903 shared the Nobel Prize for Physics with Pierre and Marie Curie, who in 1898 had isolated radium from pitchblende.

From *Nature* 13 December 1952.

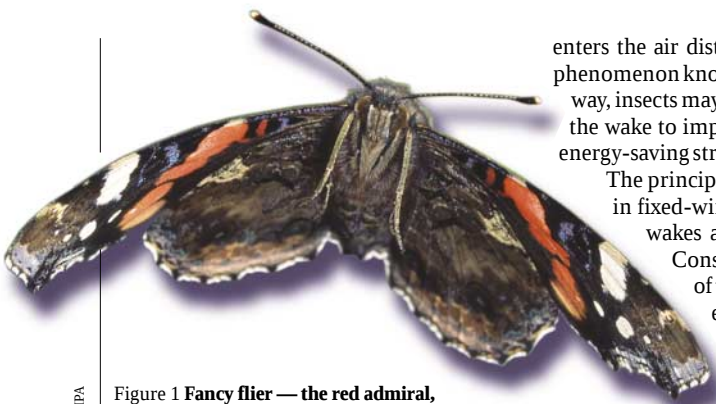


Figure 1 **Fancy flier — the red admiral, *Vanessa atalanta*.**

with superfast digital electronics.

Understanding of insect aerodynamics has to start with the wing-beat cycle of alternating downstroke and upstroke. At the beginning of the downstroke, the wing (as seen from the front of the insect) is in the uppermost and rearmost position, with the leading edge pointing forwards. The wing is then pushed downwards and forwards and rotated continuously, so that the angle of attack changes considerably during this downward motion. At the end of the downstroke, the wing is twisted rapidly, so that the leading edge points backwards, and the upstroke begins. During the upstroke, the wing is pushed upwards and backwards and rotated again, which changes the angle of attack throughout this motion. At the highest point, the wing is twisted again, so that the leading edge points forward and the next downstroke begins. So the wing is never in steady motion: it stops twice in each wing-beat cycle, and accelerates and decelerates between the stroke reversals. The resulting airflow varies considerably in time, but exhibits no turbulence (that is, it is laminar). This 'unsteady action' was first elucidated by Weis-Fogh⁹ and decisively demonstrated by Ellington¹⁰.

Weis-Fogh discovered the 'clap-and-fling' mechanism, in which the wings touch each other (clap) and then separate rapidly (fling), producing a short-lived inflow of air as vortices, which provide extra lift. Butterflies and certain wasps use clap-and-fling, but other insects do not, presumably because of the wing bashing involved with this technique. Instead, they increase lift by generating a leading-edge vortex (LEV)^{1,2}, which persists during each wing stroke and is shed at the end of it. In this process, a large vortical structure is created along the leading edge of the wing, where it persists despite the wing's acceleration and deceleration during each stroke or half-cycle.

Insects combine use of the LEV with exploitation of the wake shed from a wing's trailing edge — that is, the sheets of vortices that are left behind or, in hovering, are pushed downwards. In the hover, the wing retraces its path during each half-cycle. So it

enters the air disturbed by its own wake, a phenomenon known as wake capture. In this way, insects may recycle the momentum of the wake to improve lift, so using a subtle energy-saving strategy⁶.

The principle of LEV has been applied in fixed-wing aircraft¹¹, and unsteady wakes arise in helicopter flight¹².

Consequently, the observations of the aerodynamics involved, especially those of hovering, are amenable to mathematical modelling³. But data on flying organisms have been obtained only from tethered insects or electromechanical wing models. Tethered flight, in which an insect is glued to a metal rod and made to fly artificially, is clearly unnatural. Scaled models are experimentally more convenient. But the relevant kinematics are only approximations, and the models lack an elastic wing response and (obviously) do not respond to stimuli such as visual cues.

The problem, of course, is that free-flight data on insects are fiendishly difficult to obtain: the animals are tiny, their wings beat rapidly and techniques that do not impede movement are required. Getting detailed yet accurate quantitative data remains out of reach. But Srygley and Thomas⁴ show that, despite all the difficulties, valuable qualitative experiments can be carried out.

This is not the first time free-flight information has been gathered (it's been done with dragonflies¹³). Rather, the novelty of Srygley and Thomas's work lies in their imaging and assessment of the gross features of the airflow around their butterflies. They then interpret these features by seeking flow patterns among the smoke streamlines, using an approach known as critical point theory. This theory has a stock of typical mathematical patterns, each with known properties. If any of these patterns can be spotted in the smoke visualization images, inferences can be made about the aerodynamics involved.

For example, Srygley and Thomas could find no discernible spanwise flow (that is, air flow from wing base to tip) inside the LEV in red admirals; such a flow occurs strongly in the *Manduca* hawkmoth¹ but weakly in fruitflies². And they observe the existence of two parallel LEVs when a butterfly accelerates, but none at all when it flies forward steadily. This would suggest that the double vortices are a 'high lift' device used when rapid motion is required. Overall, however, the most remarkable finding is that red admirals (Fig. 1) seem to employ all the known lift-generating mechanisms in flight: clap-and-fling, LEVs, wake capture and rotational lift (this last is akin to the extra lift conferred on a tennis ball by spin). The butterflies appear to switch effortlessly among these mechanisms from stroke to stroke.

Disadvantages of this type of visualization

experiment are that the smoke obscures three-dimensional structures in the airflow, and the smoke streamlines become smudged. Moreover, the pattern-spotting device is the observer's eye, so interpretations are not rigorously objective. Nonetheless, Srygley and Thomas's skilful work has revealed the aeronautical versatility of the red admiral.

Butterflies are the preferred diet of some insect-eating birds, and highly manoeuvrable flight may be essential to avoid becoming easy prey⁵. The new work provides indirect support for that view. It also leaves one wondering how, with such a small brain, insects exercise flight control over such a wide range of aerobatic abilities. If engineers ever understand that, there will be a revolution in aeronautics. ■

Planetary science

Seeing double in the Kuiper belt

Daniel D. Durda

A small fraction of Kuiper-belt objects are known to be accompanied by large moons. These double worlds may have formed in the earliest days of the Solar System through comparatively gentle gravitational encounters.

Beyond the orbit of the planet Neptune, in the deep freeze of the outermost region of the Solar System, there resides a vast population of icy worlds. The Kuiper belt is the source of many of the comets that periodically pass through the inner Solar System, and it offers planetary scientists a valuable laboratory for studying the accretion of planet-forming material and for understanding the evolution of the outer Solar System¹. The formation of binary systems, objects with moons, in the Kuiper belt is the subject of much recent attention. Although a number of binaries have indeed been observed there, formation models that work for other small-body populations seem not to work as well or at all in the Kuiper belt. But on page 643 of this issue², Goldreich, Lithwick and Sari propose a new mechanism for the formation of these binaries that is backed up by observational data.

Until the discovery of the first Kuiper-belt object (KBO) in 1992, Pluto — 2,300 km in diameter and composed of a mix of ice and rock — seemed to be an anomalous interloper beyond the realm of the gas-giant planets (Fig. 1). Today, over 600 new worlds are known to reside in the Kuiper belt, along with an estimated population of some 10⁵ objects larger than 100 km in diameter. These objects now provide a natural context for Pluto as the largest member (so far) of a disk of icy minor planets, left over from the formation of the outer planets. Similarly, Pluto's large moon Charon seems a little less of an exception now that a few per cent of

Rafał Żbikowski is in the Department of Aerospace, Power and Sensors, Cranfield University, Royal Military College of Science, Shrivenham, Swindon SN6 8LA, UK.
e-mail: zbkowski@rmcs.cranfield.ac.uk

1. Ellington, C. P. *et al.* *Nature* **384**, 626–630 (1996).
2. Birch, J. M. & Dickinson, M. H. *Nature* **412**, 729–733 (2001).
3. Żbikowski, R. *Phil. Trans. R. Soc. Lond. A* **360**, 273–290 (2002).
4. Srygley, R. B. & Thomas, A. L. R. *Nature* **420**, 660–664 (2002).
5. Dudley, R. *The Biomechanics of Insect Flight: Form, Function, Evolution* (Princeton Univ. Press, 2000).
6. Dickinson, M. H. *et al.* *Science* **288**, 100–106 (2000).
7. Gullan, P. J. & Cranston, P. S. *The Insects: An Outline of Entomology* 2nd edn (Blackwell, Oxford, 2000).
8. Strausfeld, N. J. *Atlas of an Insect Brain* (Springer, Berlin, 1976).
9. Weis-Fogh, T. *J. Exp. Biol.* **59**, 169–230 (1973).
10. Ellington, C. P. *Phil. Trans. R. Soc. Lond. B* **305**, 1–181 (1984).
11. Lamar, J. E. *Prog. Aerospace Sci.* **34**, 167–217 (1998); erratum **34**, 543 (1998).
12. Leishman, J. G. *Principles of Helicopter Aerodynamics* (Cambridge Univ. Press, 2000).
13. Wakeling, J. M. & Ellington, C. P. *J. Exp. Biol.* **200**, 543–600 (1997).

KBOs have been observed to have large satellites of their own.

The first KBO satellite was discovered only last year³, and since then another six binary systems have been observed⁴ through direct imaging, using both ground-based telescopes and the Hubble Space Telescope. The total number of satellite systems now stands at eight (including Pluto–Charon), out of 618 known KBOs. In the binaries discovered so far, the two components are similar in size but are relatively far apart, separated by some 60 to 1,000 times the radius of the larger object (with the exception of Pluto–Charon, where Charon's orbital radius

is only about 17 times the radius of Pluto).

Just a decade ago, no one knew for sure whether minor planets such as KBOs and asteroids could even have satellites. Today, dozens of such systems are known in small-body populations throughout the Solar System. In addition to the diversity they bring to the bestiary of Solar System objects, the discovery and observation of KBO and asteroid satellites provides a wealth of information on the physical properties of these objects. From observations of the sizes and periods of satellite orbits, the total mass of the binaries and the individual component masses can be derived. In turn, determinations of the bulk densities of individual objects yield crucial insights into the composition and internal structure of these distant worlds.

The diverse history and evolution of minor planets and their satellites are further illuminated by models for the formation of binary systems⁵. In both the main asteroid belt (between Mars and Jupiter) and the Kuiper belt, mutual collisions play an important, if not dominant, role in shaping the properties of these objects. So mechanisms involving impacts are an obvious possibility for forming satellite systems: material ejected from cratering impacts might accrete into orbiting satellites; rotational fission of target objects might be induced by large oblique impacts; and debris from the break-up of a target body might form gravitationally bound pairs. Indeed, computer simulations of impacts between asteroids are proving successful in generating model satellite systems with properties that are consistent with those of some main-belt asteroid binaries⁶. Models for the formation of the tightly bound Pluto–Charon pair also seem to favour giant impacts similar to that thought to be responsible for the formation of our own Moon⁷.

But for the large, widely separated KBO satellites, a collisional origin is hard to understand — unless the target objects were

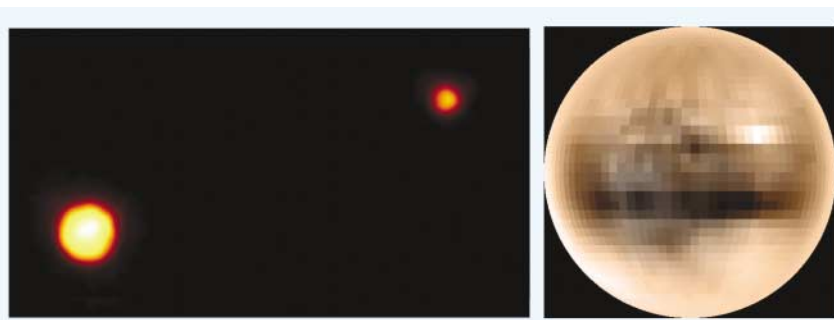


Figure 1 Pluto and its satellite Charon. The outermost planet of the Solar System resides in the Kuiper belt and is one of a number of objects there to have a satellite (left panel). How such binary systems formed in the Kuiper belt is not understood, but Goldreich *et al.*² present a model of soft gravitational interactions that could account for their existence. Pluto is the most extensively studied object in the Kuiper belt. This true-colour image from ground-based observations (right panel) suggests that the planet's surface is dominated by frozen nitrogen, with traces of methane and carbon monoxide. A NASA mission to Pluto, and the Kuiper belt, will be launched in 2006.

substantially smaller than predicted⁸. Collisions between large KBO targets and large projectiles (required to loft enough satellite-forming material into orbit) are so rare in the present-day Kuiper belt that this mechanism cannot explain the observed systems. Also, impact simulations tend to produce satellites with orbital separations of only a few times the primary's radius.

Of course, there may be many smaller, as yet undetected binary systems that were indeed formed by the same collisional mechanisms as found in the main asteroid belt. But more plausible mechanisms for the formation of large, widely separated binary pairs may involve close encounters between objects of comparable mass during primordial times, when the Kuiper belt was more populous and dynamically 'colder' than it is today.

One model⁹ proposes that under these conditions a physical collision can occur between two objects of comparable size while under the gravitational influence of a third object. The combined body resulting from that impact can remain bound to the third body, forming a satellite, if its speed after the collision is less than the speed needed to escape the gravity of the third body. The volume of space in which such collisions can occur increases with distance from the third body, so the number of satellites formed by this mechanism should increase with increasing orbital separation.

Goldreich and colleagues² propose instead that transient binaries may form when two large objects approach near enough to each other that their mutual gravitational attraction is stronger than the disruptive tidal force of the distant Sun. The transient binary is then stabilized as energy is lost from the pair through numerous encounters with passing smaller objects ('dynamical friction') or by gravitational interaction with a nearby third large body. Further loss of orbital energy through continued dynamical friction 'hardens' the binaries, tending to reduce the orbital separations. So, if Goldreich *et al.* are right, in the present-day Kuiper belt the probability of finding satellites should be greater for smaller values of orbital radius. The model further predicts that the ongoing Hubble Space Telescope survey for Kuiper-belt binaries should find that about 5% of KBOs are accompanied by large satellites — and this is indeed roughly compatible with observations made to date.

Despite the seeming conflict between the predictions of these two models, and the fact that there remain some crucial details of timing to be worked out, both models offer plausible mechanisms for binary formation in the Kuiper belt, which may even have acted in concert in primordial times. As new KBO binaries are discovered, their properties — particularly the statistics of their orbital

separations — will provide important constraints on these models and others. A spacecraft mission to Pluto–Charon and the Kuiper belt will provide wonderfully detailed information on the physical properties of at least one KBO binary and remains a top priority in outer Solar System exploration¹⁰. The exciting prospect of testing competing binary formation models with ongoing observations exemplifies a healthy state of affairs in exploring the frontiers of the Solar System.

Daniel D. Durda is in the Department of Space Studies, Southwest Research Institute, Boulder, Colorado 80302, USA.

e-mail: durda@boulder.swri.edu

1. Farinella, P., Davis, D. R. & Stern, S. A. in *Protostars and Planets IV* (eds Mannings, V., Boss, A. P. & Russell, S. S.) 1255–1282 (Univ. Arizona Press, Tucson, 2002).
2. Goldreich, P., Lithwick, Y. & Sari, R. *Nature* **420**, 643–646 (2002).
3. Veillet, C. *et al.* *Nature* **416**, 711–713 (2002).
4. www.boulder.swri.edu/ekonews/objects/binaries.html
5. Merline, W. J. *et al.* in *Asteroids III* (eds Bottke, W. F., Cellino, A., Paolicchi, P. & Binzel, R. P.) 289–312 (Univ. Arizona Press, Tucson, 2002).
6. Durda, D. D., Bottke, W. F., Asphaug, E. & Richardson, D. C. *Bull. Am. Astron. Soc.* **33**, 1134 (2001).
7. Canup, R. M. & Asphaug, E. *Nature* **412**, 708–712 (2001).
8. Stern, S. A. *Astron. J.* **124**, 2300–2304 (2002).
9. Weidenschilling, S. J. *Icarus* **160**, 212–215 (2002).
10. National Research Council *New Frontiers in the Solar System: An Integrated Exploration Strategy* (2002); www.nationalacademies.org/ssb/newfrontiersfront.html

Physiology

Is brain sympathetic to bone?

Jeffrey S. Flier

The fat-derived hormone leptin is best known for its effects on weight. But it also influences bone density, and new work reveals a role for the sympathetic nervous system in mediating this effect.

It seems intuitive that the various aspects of our physiology do not function in isolation but instead interact, often in complex ways. For instance, the observation that overweight people are somewhat less susceptible to osteoporosis (reduced bone mass) hints at a link between obesity and the skeletal system. Yet the basis for such interactions is often unclear. It was first suggested that the

effects of added weight-bearing might somehow protect bone mass. Later experimental studies¹, however, implicated leptin. This hormone, which is produced by the body's fat cells, was already known to interact with neurons in the brain² and thereby influence weight — for instance, by reducing appetite and increasing energy expenditure^{3,4}. It was then discovered¹ that, in mice, leptin is also

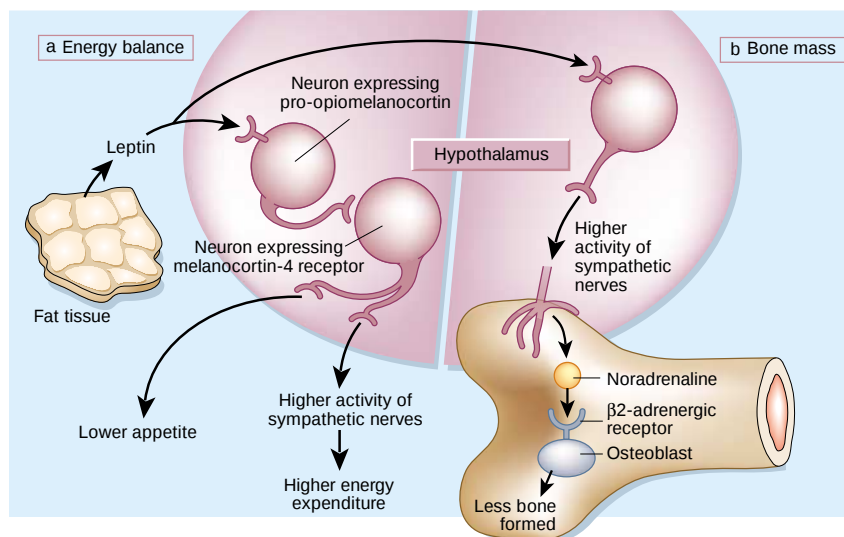


Figure 1 Molecular connections between fat, appetite, energy expenditure and bone mass. Leptin is a hormone that is released from fat cells in proportion to body fat, and travels in the blood to the brain. **a**, Leptin acts on neurons in the hypothalamus, including (but not only) those depicted here, to regulate appetite and energy expenditure. **b**, Takeda *et al.*⁵ find that leptin also acts on hypothalamic neurons (although their identity is unknown) to regulate bone mass. They propose that this stimulates the activity of sympathetic (involuntary) nerves that penetrate the bone, there releasing noradrenaline. This neurotransmitter in turn binds to the β 2-adrenergic receptors on bone-forming cells (osteoblasts), inhibiting their activity. So leptin reduces bone mass. Perhaps contrary to expectations, obesity offers some protection from reduced bone mass. It has been speculated that this is due to resistance to leptin's bone-reducing effects.

antiosteogenic (it reduces bone mass). The finding led to speculation that increased bone mass in obese people might result not from an adaptive response to the stress of added pounds, but from resistance to leptin's antiosteogenic activity, just as resistance to leptin's appetite-reducing effects characterizes most obesity⁴. Writing in *Cell*, Takeda *et al.*⁵ have delved deeper into the way in which leptin reduces bone mass under normal circumstances.

Leptin is released into the bloodstream in proportion to the amount of body fat. It regulates the body's energy balance — the difference between food intake and energy expenditure — by binding to certain receptor proteins, which are expressed on, and regulate the activity of, specific neurons in the hypothalamus of the brain (Fig. 1a). These molecular and neuronal targets of leptin were identified through a combination of genetic, pharmacological and neuroanatomical approaches. Takeda *et al.*⁵ have now applied the same range of techniques to look at how leptin affects bone. Their first, remarkable conclusion is that, although leptin influences both energy balance and bone mass by acting on the hypothalamus, the two processes involve different proteins and neurons.

For example, many of leptin's effects on appetite and energy expenditure involve the activation of neurons found in one region of the hypothalamus, the arcuate nucleus. These neurons produce the protein pro-opiomelanocortin, the precursor of α -melanocyte-stimulating hormone. When this hormone is released from the neurons, it binds to the melanocortin-4 receptor on other neurons, thereby activating them and leading to effects on energy balance⁶. When

these receptors are absent or blocked in mice, the animals become severely obese and no longer respond to the energy-regulating effects of leptin⁷. Yet, as Takeda *et al.* show, these mutants apparently still respond to leptin's bone-density-reducing effects, as their bone mass does not increase. Similarly, when the authors damaged these arcuate neurons by administering monosodium glutamate, leptin — injected into the brain — no longer affected energy balance, but still reduced bone density. This convincingly demonstrates that the weight- and bone-regulating neuronal pathways are different.

It is still not known which leptin-responsive neurons regulate bone, although one clue came from an earlier study of neuropeptide Y and the Y2 receptor protein⁸. Neuropeptide Y was known to be secreted by certain neurons in the arcuate nucleus (a process that is inhibited by various hormonal regulators, including leptin). It then binds to the Y1–Y5 receptors on other neurons, thereby regulating feeding and the activity of the involuntary nervous system. The study in question⁸ found that Y2 receptors were also involved in regulating bone synthesis and mass, because mice lacking these hypothalamic receptors had increased bone mass. This suggested a role for involuntary output from the hypothalamus, mediated through nerves on which the Y2 receptor is expressed, in regulating bone.

In theory, this involuntary output could trigger the release of other hormones that reach bone through the blood circulation, or it could affect bone directly. Takeda *et al.* have tackled this question. First, they propose that no circulating hormone (besides leptin, that is) is involved. Next, they use

various approaches to demonstrate a role for the sympathetic nervous system — a major arm of the involuntary nervous system — in controlling bone density. It was known that leptin activates sympathetic-nerve output in rodents⁹, but it was not suspected that bone biology was also affected by this.

Takeda *et al.* now propose the following model (Fig. 1b). Leptin activates hypothalamic nerves (identity unknown), which in turn activate sympathetic nerves. These extend into the bone, where they stimulate release of the neurotransmitter noradrenaline, which then stimulates β 2-adrenergic receptors on osteoblasts (bone-forming cells), inhibiting osteoblast activity.

The authors invoke several lines of evidence to support this model. First, dopamine- β -hydroxylase-deficient mice, which cannot synthesize noradrenaline¹⁰, have high bone density, presumably because the bone-forming cells remain active. Although such mice respond to leptin by reducing their fat mass substantially, there is no effect on bone. Second, leptin-deficient mice show a reduced rate of firing of sympathetic nerves and high bone mass; drugs that stimulate β -adrenergic receptors do not affect the body fat of these mice, but do cause severe loss of bone — apparently by reducing the rate of bone formation by osteoblasts. Finally, and perhaps most surprisingly, when mice are given the β -adrenergic-receptor blocker propranolol, which is commonly used to treat cardiovascular ailments, bone mass is increased. This occurs even in mice with depleted oestrogen levels following removal of the ovaries, which otherwise causes bone mass to plummet.

How readily can we extrapolate these

Biomechanics

Frogs in and out of phase

Many animals use several different gaits — patterns of locomotion that change abruptly at a critical speed. People walk to go slowly and run to go fast. Horses and other quadrupedal mammals walk, trot and gallop. In flight, birds and bats have slow and fast gaits. Many fish swim slowly with their fins and fast with their whole bodies. On land, frogs crawl slowly and hop faster.

Now we learn from Sandra Nauwelaerts and Peter Aerts that frogs also change gait when they swim (*J. Zool.* **258**, 183–188; 2002). Frogs swim by kicking water backwards with their webbed feet. Usually they kick with both hind legs simultaneously (in-phase swimming), but Nauwelaerts and Aerts find that in slow swimming the hind legs move alternately (out-of-phase swimming). The subject of their attention, the species *Rana esculenta*, is shown here in static pose.

Measurements of oxygen consumption show

that humans and horses change gait at speeds at which the slower gait becomes less energy-efficient than the faster one. The same may be true for flying birds. Fish, however, seem to change gait because their fin muscles cannot provide enough power for high speeds. Do swimming frogs save energy by changing gait, or do they have to change gait to develop enough power to swim fast?

From measurements on video sequences, Nauwelaerts and Aerts calculate that out-of-phase swimming uses less power than in-phase swimming at the same speed. During in-phase swimming, the animal accelerates by kicking with both feet, then decelerates as it brings its feet forward for the next kick. During out-of-phase swimming, the frog's speed fluctuates much less because each leg is brought forward while the other is kicking. Hydrodynamic drag is approximately proportional to the square of the speed, so speed



fluctuations increase the mean drag.

This argument suggests that out-of-phase swimming should be the more economical gait at all speeds. However, it may be less effective at generating thrust, because the feet interact when they are used together. So frogs may be obliged to use the less economical gait to generate enough thrust to swim fast.

R. McNeill Alexander

R. McNeill Alexander is at the School of Biology, University of Leeds, Leeds LS2 9JT, UK.
e-mail: r.m.alexander@leeds.ac.uk

insights from mouse models to humans? Several facts call for caution until more data are available. First, humans with anorexia nervosa lack both leptin and oestrogen; yet anorexic patients rapidly lose bone density¹¹. The model proposed by Takeda *et al.* would not predict this; indeed, the authors found that mice lacking both leptin and oestrogen have increased bone mass. Second, humans with congenital leptin deficiency are reported¹² to have normal, rather than raised, bone density, which does not fall after leptin-replacement therapy, in contrast to the findings of Takeda *et al.* in leptin-deficient mice. Similarly, whereas Takeda *et al.* found that mice lacking melanocortin-4 receptors have normal bone density, humans without these receptors have increased bone density¹³. It remains to be seen whether these discrepancies represent species differences or result from differences in the methods used to quantify bone mass.

Finally, Takeda and colleagues' results would seem to suggest that the millions of patients who have been treated with 'beta blockers' (such as propranolol) for hypertension and other cardiovascular disorders should have increased bone density. But

that fact seems to have escaped the research and clinical communities. Despite such inconsistencies, however, the implications of the new results for human bone biology and for drug development are enormous. It is time for clinical investigators to quickly fill in the blanks.

Jeffrey S. Flier is at the Beth Israel Deaconess Medical Center, Research and Academic Affairs, 330 Brookline Avenue, Boston, Massachusetts 02215, USA.

e-mail: jflier@caregroup.harvard.edu

1. Ducey, P. *et al.* *Cell* **100**, 197–207 (2000).
2. Elmquist, J. K., Maratos-Flier, E., Saper, C. B. & Flier, J. S. *Nature Neurosci.* **1**, 445–450 (1998).
3. Zhang, Y. *et al.* *Nature* **372**, 425–432 (1994); erratum **374**, 479 (1995).
4. Ahima, R. S. & Flier, J. S. *Annu. Rev. Physiol.* **62**, 413–437 (2000).
5. Takeda, S. *et al.* *Cell* **111**, 305–317 (2002).
6. Cowley, M. A. *et al.* *Nature* **411**, 480–484 (2001).
7. Huszar, D. *et al.* *Cell* **88**, 131–141 (1997).
8. Baldock, P. A. *et al.* *J. Clin. Invest.* **109**, 915–921 (2002).
9. Haynes, W. G., Morgan, D. A., Walsh, S. A., Mark, A. L. & Sivitz, W. L. *J. Clin. Invest.* **100**, 270–278 (1997).
10. Alaniz, R. C. *et al.* *Proc. Natl Acad. Sci. USA* **96**, 2274–2278 (1999).
11. Muñoz, M. T. & Argente, J. *Eur. J. Endocrinol.* **147**, 275–286 (2002).
12. Farooqi, I. S. *et al.* *J. Clin. Invest.* **110**, 1093–1103 (2002).
13. Farooqi, I. S. *et al.* *J. Clin. Invest.* **106**, 271–279 (2000).

Oceanography

Gas hydrates on the brink

Ingo A. Pecher

Huge amounts of methane are locked up in deposits that lie deep beneath the sea floor. New seismic images reveal that these deposits possess unexpected features that might affect their stability.

Gas hydrates are an ice-like form of water that has cavities containing gas — usually methane. They exist in vast quantities beneath the ocean floor in certain areas, especially continental margins, where the methane is generated mostly from the bacterial breakdown of organic matter¹, and they remain stable under seafloor conditions that are typical of water depths of more than 300 m or so². There is considerable interest in these hydrates, largely because of their potential as an energy resource and because of the effects that released methane may have, or have had, on climate. Methane release will depend on gas-hydrate stability. The zone of stability extends downwards into the sea-floor sediments until, because of increasing temperature, the 'base of gas hydrate stability' (BGHS) is reached some tens to hundreds of metres beneath the sea floor.

In their report on page 656 of this issue, Wood *et al.*³ present new data that dramatically change our perception of the BGHS. From conventional seismic data, acquired using sources and receivers towed close to the surface, the BGHS appears to be a continuous boundary that runs more or less parallel

to the sea floor. But the images presented by Wood *et al.*, from survey work at locations off the east and west coasts of North America, show that the BGHS exhibits considerable 'roughness'. These features, which may well be a general characteristic of the BGHS elsewhere in the world, could greatly affect the stability of gas-hydrate reservoirs.

The new data were acquired with a seismic source and an array of receivers towed close to the sea floor. This configuration provides an increase in resolution of almost one order of magnitude over previous methods. The images reveal small but densely packed pockets of gas at the BGHS which, in conventional data, appear to be part of continuous gas layers. More intriguingly, narrow, near-vertical 'wipeout zones' (opaque zones with little reflected seismic energy) puncture the gas-hydrate-bearing sediments. Wood *et al.* interpret these wipeout zones as gas chimneys (Fig. 1a), through which methane can migrate upwards, probably along faults, well into the regional zone of gas-hydrate stability.

Wood *et al.*³ support this interpretation by modelling fluid and heat flow through the

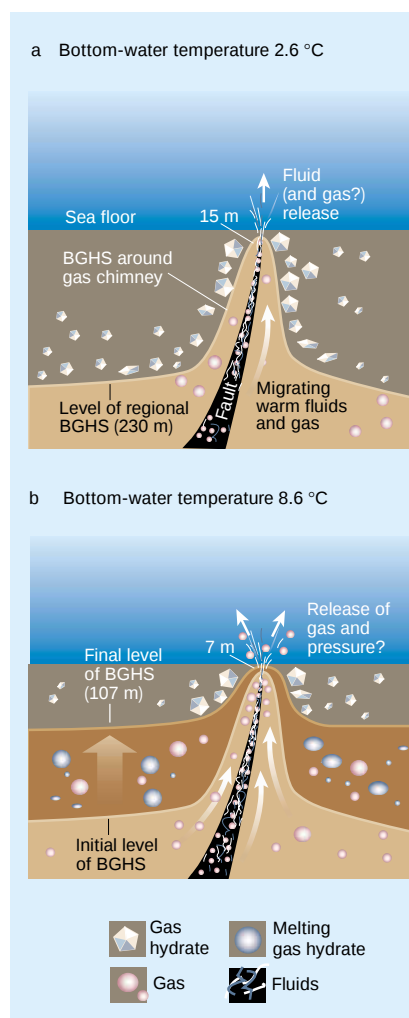


Figure 1 Gas chimneys and their possible effects on gas-hydrate stability. a, Wood *et al.*³ interpret their data as revealing these chimneys, which enlarge the area of the 'base of gas hydrate stability' (BGHS) and mean that more gas hydrates are located near the stability boundary. The situation depicted here is for a water depth of 1,260 m; a bottom-water temperature of 2.6 °C; and the top of the gas chimney at 15 m depth below the sea floor. b, Estimation of the effect of a sudden (in geological terms) increase of bottom-water temperature¹⁰ by 6 °C. From calculations based on ref. 11, a thermal pulse travelling through the sediment would reach 15 m depth after only about 55 years, and 230 m after 14,000 years. The existence of gas chimneys surrounded by highly concentrated gas hydrates close to the sea floor should significantly reduce the lag between an increase of bottom-water temperature and the onset of gas-hydrate dissociation. Above the chimney, the BGHS moves up to 7 m below the sea floor and the regional BGHS to 107 m, causing gas release and a build-up of pressure. Speculatively, the chimneys might act as valves by allowing gas migration through the overlying sediment, and into the ocean, so relieving the pressure.

faults. The results suggest that warm fluids moving relatively quickly through highly permeable faults can keep the faults and surrounding sediments warm enough to prevent gas-hydrate formation. That is, although on a regional scale the faults lie well within the zone of gas-hydrate stability, locally, on a scale of several metres, conditions are such that methane will not be transformed to hydrate. Based on evidence of small lenses with high gas-hydrate concentrations⁴, it has been proposed that such gas chimneys existed in the geological past. These lenses may have formed from gas that was injected along chimneys into the regional gas-hydrate zone.

The chimneys increase the amount of hydrates located close to the BGHS by enlarging the area of the boundary by a factor of 0.5–3. So the gas-hydrate reservoir may be more susceptible to changes in environmental conditions, such as variations in sea level or water temperature, than previously thought. Near the chimneys, high gas-hydrate concentrations would be expected to occur close to the sea floor because of a high methane supply. A sudden change of bottom-water temperature, which is likely to be the most significant environmental cause of gas-hydrate dissociation⁵, would affect hydrates there more directly and faster than it would deeply buried deposits.

The effects of the chimneys on methane release might go further. Gas-hydrate dissociation in response to environmental change is predicted to take place primarily near the BGHS. As long as gas hydrates remain stable at the sea floor, upward-migrating gas would be expected to become 'trapped' by hydrates before reaching the ocean. So most explanations⁶ of large-scale methane release into the ocean invoke catastrophic seafloor failure, in which pressure build-up from gas generated at the BGHS produces an underwater eruption. It has long been realized that gas may migrate upwards along faults in a gas-hydrate zone⁷. But it was not clear why the gas would not be transformed to hydrate within the faults.

From their modelling study, Wood *et al.*³ predict that in the vicinity of faults the BGHS will almost reach the sea floor. So, to speculate, if the fluids within the faults remain warm up to the sea floor, methane gas could migrate the last few metres through the sediments to the ocean. The chimneys may therefore provide methane migration pathways from the BGHS through the gas-hydrate zone, and (for example, following bottom-water warming) may act as valves by steadily or episodically releasing methane into the ocean without seafloor collapse (Fig. 1b). A similar mechanism has been proposed⁸ to be occurring at the Blake Ridge Depression, off South Carolina, which has been interpreted as a large hydrate-bearing sediment 'wavefield' (the underwater analogy

of sand dunes). Fluid- and gas-migration paths cutting through the gas-hydrate stability zone have been detected where individual sediment 'dunes' meet, and gas may be released episodically through these paths into the ocean.

The study by Wood *et al.* adds considerably to our knowledge of methane transfer from gas hydrates into the ocean. But what about the possible effects on climate? Methane has a much higher greenhouse potential than CO₂ — molecule for molecule, it is estimated to be 3.7 times more powerful in its warming effect⁹. Carbon-isotope anomalies in oceanic sediments have been linked to a release of methane from gas hydrates¹⁰, one implication being that such releases might be a cause of warming episodes in Earth's history. It is unclear, however, how much of this methane reaches the atmosphere — most of it may be oxidized to CO₂ in the ocean¹⁰. So before we can quantify any connection between

gas hydrates and climate change, we need a better understanding of what happens to methane in the ocean. ■

Ingo A. Pecher is at the Institute of Geological and Nuclear Sciences Ltd, PO Box 30 368, Lower Hutt, Wellington, New Zealand.

e-mail: i.pecher@gns.cri.nz

1. Wellsbury, P. & Parkes, J. R. in *Natural Gas Hydrate in Oceanic and Permafrost Environments* (ed. Max, M. D.) 91–104 (Kluwer, Dordrecht, 2000).
2. Sloan, E. D. *Clathrate Hydrates of Natural Gases* (Dekker, New York, 1990).
3. Wood, W. T., Gettrust, J. F., Chapman, N. R., Spence, G. D. & Hyndman, R. D. *Nature* **420**, 656–660 (2002).
4. Gorman, A. R. *et al.* *Geology* **30**, 327–330 (2002).
5. Dickens, G. R. *Org. Geochem.* **32**, 1179–1193 (2001).
6. Mienert, J., Posewang, J. & Baumann, M. in *Gas Hydrates: Relevance to World Margin Stability and Climatic Change* (eds Henriet, J. P. & Mienert, J.) 275–291 (Geol. Soc., London, 1998).
7. Dillon, W. P. *et al.* in *Proc. 29th Offshore Technol. Conf.* 201–209 (Offshore Tech. Conf., Dallas, 1997).
8. Holbrook, W. S. *et al.* *Geology* **30**, 467–470 (2002).
9. Lashof, D. A. & Ahuja, D. R. *Nature* **344**, 529–531 (1990).
10. Katz, M. E., Pak, D. K., Dickens, G. R. & Miller, K. G. *Science* **286**, 1531–1533 (1999).
11. Fowler, C. M. R. *The Solid Earth* (Cambridge Univ. Press, 1990).

HIV

Conformational camouflage

Theodore Jardetzky

A study of the thermodynamics of antibody binding to a crucial HIV protein has shed light on why the virus so effectively evades the antibody arm of the immune response. Changes in the protein's conformation are the key.

Despite two decades of research, HIV has defied immunologists' best efforts to develop a broadly protective vaccine. Writing on page 678 of this issue, Kwong and colleagues¹ suggest a new explanation. The authors reveal an unusual mechanism by which HIV might evade human antibodies, whether they are generated in response to a vaccine or to the viral infection itself.

HIV infects T cells of the immune system, leading to their depletion and the associated loss of immune function that is characteristic of AIDS. The viral glycoproteins gp120 and gp41 mediate infection by forming an oligomeric complex on the HIV surface, comprising three gp120 and three gp41 molecules, that targets the virus to T cells. The gp120 protein binds to two receptors on T cells, CD4 and a member of the chemokine-receptor family (Fig. 1a, overleaf), triggering the fusion of viral and cellular membranes and viral entry into the cell. Receptor binding is thought to induce conformational changes in gp120 that allow gp41 to drive membrane fusion by also undergoing a large structural change². During an infection, soluble monomeric gp120 is also produced.

One arm of the immune response to intruders is the production, by B cells, of antibodies that recognize key features of the invaders, resulting in their neutralization or destruction. For instance, HIV's gp120 pro-

tein is a major target for antibodies. This fact, and the knowledge that gp120 is the major viral surface glycoprotein and a key mediator of the entry process, would seem to suggest that gp120 is an ideal target for vaccines to stimulate anti-HIV immunity. Unfortunately, though, the virus has several strategies for evading recognition by the immune system, hampering attempts to produce a vaccine^{3–5}. For instance, HIV mutates rapidly during replication, producing many gp120 variants and an HIV genome that can vary greatly both within a single person and across larger geographical populations⁴. So immunizations with monomeric gp120 have generally been disappointing, yielding antibodies that are too specific to protect against the broad range of HIV variants⁵. Yet despite all of this variation, the overall structure of gp120 must be at least partially conserved for the protein to function in receptor binding and cellular entry. Theoretically, the receptor-binding sites could provide a target for broadly neutralizing antibodies — so why haven't they done so? Kwong *et al.*¹ suggest an answer.

The structure of monomeric gp120, solved several years ago⁶, provides a three-dimensional template for mapping the regions that bind gp41, the receptors and antibodies. The core of gp120 consists of two structural domains with an intervening

Rotting smell of dead-horse arum florets

These blooms chemically fool flies into pollinating them.

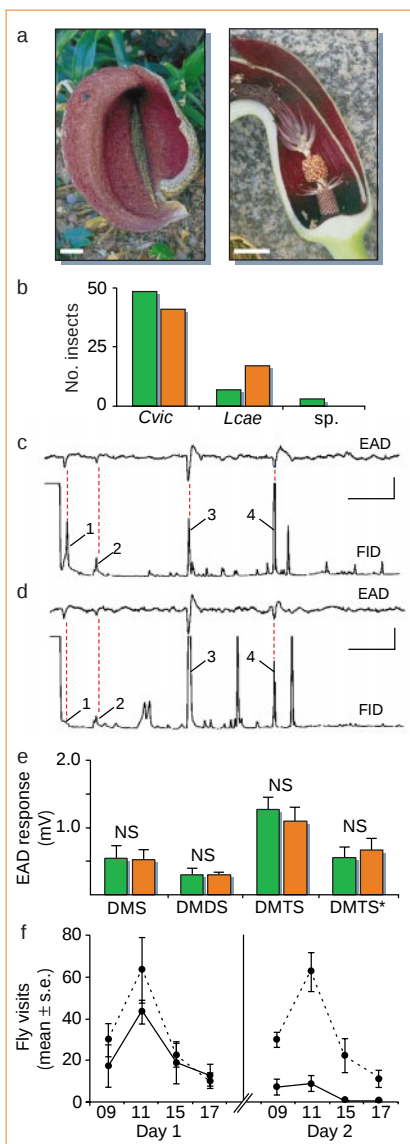


Figure 1 Olfactory mimicry by the dead-horse arum (*Helicodiceros muscivorus*). **a**, Inflorescence (left), and a transverse section (right) showing receptive female (bottom) and male (top) florets separated by spines. Scale bars, 5 cm. **b**, Distribution of insects visiting flowering arum plants (green bars; $n=6$) and carcasses (orange bars; dead gulls, $n=3$). Females comprise 78% and 100% of the arum and carrion catch, respectively. Species: Cvic, *Calliphora vicina*; Lcae, *Lucilia caesar*; sp., other Calliphoridae species. **c**, **d**, Antennal detection of arum (**c**) and carcass (**d**) volatiles. FID, flame ionization detection; EAD, electro-antennographic detection; dashed lines refer to EAD. Horizontal scale bar, 1 min; vertical bar (for EAD trace), 1 mV. EAD active peaks were identified as follows: 1, dimethyl sulphide (DMS); 2, dimethyl disulphide (DMDS); 3 and 4, two conformers of dimethyl trisulphide (DMTS and DMTS*). **e**, Mean ($\pm$ s.e.) EAD responses (green, arum; orange, carrion) elicited by the oligosulphides from eight GC-EAD runs. NS, no significant difference (paired t -test). **f**, Fly visits during the 2-day flowering period in untreated plants and in plants treated artificially with oligosulphides during the second day (dashed line), showing that these restore attractivity.

Deceit by resource mimicry has evolved as a pollination strategy in several plant species^{1–3} and is particularly elaborate in a plant known as dead-horse arum (*Helicodiceros muscivorus*; Araceae: Aroideae), which may fool flies into pollinating it by emitting a smell like a dead animal — an important oviposition resource for these insects. Here we confirm that the composition of volatiles from these flowers and from a rotting carcass is strikingly similar and show that the pollinators respond in the same way to chemicals from both sources. This remarkably complex mimicry must have evolved to exploit insects as unrewarded pollinators.

The dead-horse arum grows on small islands off the coasts of Sardinia, Corsica and the Balearics^{4,5}. For pollination, it lures insects into a trap chamber that surrounds the female florets (Fig. 1a), where their exit is hindered by spines and filaments: insects carrying pollen from another plant thus fertilize the receptive female florets. The chamber remains closed until the male florets at the entrance (Fig. 1a) begin to produce pollen and the female florets are no longer receptive; insects that leave the chamber are coated with fresh pollen as they brush past the male florets. The pollinators are primarily females of two Calliphoridae blowfly species, which are also carrion visitors (Fig. 1b).

We have investigated whether the carrion odour of the dead-horse arum represents a case of genuine mimicry. If it does, the chemical composition of the arum odour should be similar to that from a carcass; moreover, both odours should elicit the same olfactory response from blowflies if identical channels are used to convey olfactory information.

To identify the volatile compounds associated with the two different sources, we used gas chromatography with simultaneous flame ionization detection and electro-antennographic detection⁶ to characterize the odour of samples from the arum (Fig. 1c) and from a carcass (Fig. 1d). Calliphoridae blowfly antennae were stimulated with each odour, both of which generated identical antennal response patterns (Fig. 1e).

The chemicals that elicit these antennal responses were identified as three structurally similar oligosulphides: dimethyl mono-, di- and trisulphide. The trisulphide derivative was present as two structural conformers in both the arum and carcass volatiles. These odorants have previously been identified in the headspace of the arum plant⁷, and oligosulphides are produced

by decaying meat during protein decomposition⁸, which should help to guide blowflies to suitable oviposition sites. The identical antennal responses to arum and carcass odours indicate that a carrion fly cannot rely on odours alone to separate the mimic from the model.

Arum flowers typically remain open for two days. The foul smell due to the oligosulphides is produced only during the first day of flowering and is not detectable in odours sampled on the second day. We recorded the number of flies that visited arum plants ($n=7$) during the two-day flowering period, and found that plants were attractive to blowflies mainly during the first day, whereas the odourless state during the second day attracted significantly fewer flies (paired t -test, $P=0.03$; Fig. 1f).

To confirm the influence of the oligosulphide compounds on the flies' behaviour, we impregnated dental cotton rolls with 100 μ l of a 1:1:1 mix of mono-, di- and trisulphide derivatives, placed them inside the trap chambers of arum plants ($n=6$) during their second, odourless day of flowering, and recorded the number of fly visits starting from day 1 (natural odour) through to the end of day 2 (artificial odour). The plants attracted a comparable number of flies on both days (paired t -test, $P=0.97$; Fig. 1f). This indicates that the oligosulphides are important components in the mimicry and that their odours attract flies to the flower.

The oligosulphide odours are also crucial for the blowflies, serving to attract them to carrion resources. The arum system can be compared to the sexual deception used by *Ophrys* orchids for pollination purposes^{9,10}; in both cases, the plants mimic odours that the insects cannot afford to ignore. The dead-horse arum flower may also use visual and thermal cues to reinforce the deception. This plant is a striking example of evolutionary cunning that exploits insects for pollination purposes.

Marcus C. Stensmyr*, **Isabella Urrut†**, **Ignazio Collu†**, **Malin Celander***, **Bill S. Hansson***, **Anna-Maria Angioy†**

*Department of Crop Science, Swedish University of Agricultural Sciences, PO Box 44, 23053 Alnarp, Sweden

e-mail: bill.hansson@vv.slu.se

†Department of Experimental Biology, University of Cagliari, 4500 Monserrato, Italy

- Linnaeus, C. Carl Linnaei Öländska och Gothländska Resa på Rikens Höglöflige Ständers Befallning Förrättad åhr 1741. Med Anmärkningar uti Oeconomien, Natural-historien, Antiquiteter och med åtskillige Figurer (Kiesewetter, Stockholm, 1745).
- Dafni, A. *Annu. Rev. Ecol. Syst.* **15**, 259–781 (1984).
- Roy, B. A. & Widmer, A. *Trends Plant Sci.* **4**, 325–330 (1990).

4. Fridlender, A. *Webbia* **55**, 7–35 (2000).
5. Rosello, J. A. & Saez, L. *Acta Bot. Barc.* **44**, 169–174 (1997).
6. Arn, H. *et al.* *Naturforsch.* **30**, 722–725 (1975).
7. Kite, G. C. *Kew Bull.* **55**, 237–240 (2000).

8. Brown, M. H. *Meat Microbiology* (Appl. Sci., London, 1982).
 9. Schiestl, F. P. *et al.* *Nature* **399**, 421–422 (1999).
 10. Schiestl, F. P. *et al.* *J. Comp. Physiol. A* **186**, 567–574 (2000).
- Competing financial interests:** declared none.

COMMUNICATIONS ARISING

Ontogenetic growth

Modelling universality and scaling

Understanding the allocation of metabolic energy between the sustenance of an organism and its growth is an important issue in ecology. West *et al.*¹ have built on their earlier attempts to explain the exponent for the allometric scaling of metabolism and have derived a single, parameterless universal curve that describes the growth of many species. Here we show that the universal curve arises from general considerations that are independent of the specific allometric model used by West *et al.* and that the data do not distinguish between a 3/4 or 2/3 exponent in the relationship between metabolic rate and mass scaling.

von Bertalanffy² (see his equation (5)) considered a general equation of the form $dm/dt = a_m m^\alpha - b_m m^\beta$, where m is the body's mass at time t , and α and β are unspecified exponents, and a_m and b_m are positive coefficients. The equation considered by West *et al.*¹ is a special case, with $\alpha = 3/4$ and $\beta = 1$, whereas von Bertalanffy studied the case with $\alpha = 2/3$ and $\beta = 1$. Such an equation can be cast in a scaling form: $dm/dt = m^\alpha f(m/M)$, where M provides a scale for the organism's

mass and is usually chosen to be the asymptotic maximum body size. The scaling function f approaches a constant value for a small argument — when $m(t)$ is small compared with M , relatively little energy is required to sustain the organism and virtually all of the metabolic energy is funnelled into growth processes. The requirement that $dm/dt = 0$ when m reaches the value M leads to the condition that $f(1) = 0$.

Moreover, the initial condition of the mass at $t = 0$ being equal to the birth mass, m_0 , must be satisfied. For von Bertalanffy's equation², $f(x) = a_\alpha(1 - x^{1-\alpha})$, with $a_\alpha = b_\alpha M^{1-\alpha}$.

We have analysed the data of West *et al.*¹ on the cow, hen and guppy to assess whether it is possible to discriminate between the two choices for α . We find that their equation (5) (which is a special case of equation (6) of ref. 2), written in the form

$$(m/M)^{1-\alpha} = 1 - [1 - (m_0/M)^{1-\alpha}]e^{-\gamma_\alpha t} \quad (1)$$

or, equivalently, as

$$r = 1 - e^{-\tau} \quad (2)$$

(with the dimensionless mass ratio $r = (m/M)^{1-\alpha}$ and the dimensionless time $\tau = -\ln[1 - (m_0/M)^{1-\alpha}]e^{-\gamma_\alpha t}$), fits the observations equally well for both values of α (Fig. 1).

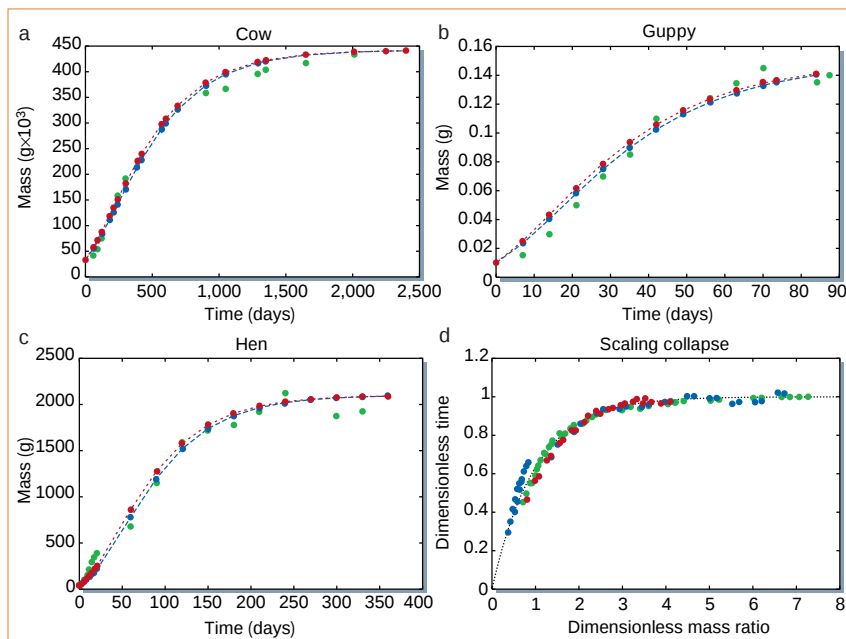


Figure 1 Growth curves for three different organisms and the collapse of mass scaling. **a–c**, Growth curves for **a**, cow; **b**, guppy; and **c**, chicken. Green, empirical data; blue, best fit obtained by West *et al.*¹ (that is, $\alpha = 3/4$); red, plot of equation (1), with $\alpha = 2/3$ and the values of M and γ as obtained in ref. 1. **d**, Scaling collapse. The universal growth curve (equation (2); dotted line) is derived from data from the three species (green, cow; blue, hen; red, guppy) for both $\alpha = 3/4$ and $\alpha = 2/3$.

For simplicity, we have chosen M from Table 1 of ref. 1. We determined $\gamma_\alpha = a_\alpha(1 - \alpha)/M^{1-\alpha}$ using the values of $a_{3/4}$ from the same table and $a_{2/3}$ for the cow, hen and guppy to be 0.62, 0.67 and 0.064, respectively, to ensure that $\gamma_{2/3} = \gamma_{3/4}$ for simplicity. (We have not tried to adjust the value of M and γ to achieve a better fit because that is beside the point here.) Also, an equally good fit of the universal curve is obtained (Fig. 1d) for the three species with both values of α . Furthermore, the form of the universal curve (equation (2)) is independent of the value of α . Thus, the existence of a universal curve indicates nothing about the value of α .

Jayanth R. Banavar*, **John Damuth†**, **Amos Maritan‡**, **Andrea Rinaldo§**

*Department of Physics, 104 Davey Laboratory, Pennsylvania State University, University Park, Pennsylvania 16802, USA
e-mail: banavar@psu.edu

†Department of Ecology, Evolution and Marine Biology, University of California, Santa Barbara, California 93106, USA

‡International School for Advanced Studies, 34014 Trieste, and INFN and the Abdus Salam International Center for Theoretical Physics, 34014 Trieste, Italy

§Dipartimento di Ingegneria Idraulica, Marittima e Geotecnica, Università di Padova, 35100 Padova, Italy

1. West, G. B., Brown, J. H. & Enquist, B. J. *Nature* **413**, 628–631 (2001).
2. von Bertalanffy, L. Q. *Rev. Biol.* **32**, 217–231 (1957).

West *et al.* reply — Of the equations that have been used to describe ontogenetic growth in terms of the rate of increase in mass, m , as a function of time, t , most are merely statistical descriptions with no mechanistic basis. Our model¹ is derived from fundamental biological and physical principles and relates growth to metabolic power at the cellular level. It is based on the allocation of resources to the maintenance and replacement of existing tissue and the production of new tissue, with the whole-body metabolic rate $B = N_c B_c + E_c(dN_c/dt)$, where B_c is the cellular metabolic rate, E_c is the energy needed to create a cell, and N_c is the total number of cells. As $m = N_c m_c$, where m_c is the average cell mass, this gives

$$dm/dt = am^\alpha - bm^\beta \quad (1)$$

where $a = B_0 m_c/E_c$, $b = B_c/E_c$, $\beta = 1$ and α is the allometric exponent for B ($\equiv B_0 m^\alpha$), taken to be 3/4 in accordance with a large body of data and with theoretical arguments^{2,3}. The asymptotic mass, for which $dm/dt = 0$, is predicted to be $M = (B_0 m_c/B_c)^4$.

Equation (1) reflects the diminishing capacity of fractal-like distribution networks to supply resources as body size increases. It has no 'arbitrary' parameters; all exponents and coefficients are derived from measured

fundamental quantities that are not directly related to growth, such as m_c , E_c , B_c and B_0 .

Equations of the form (1) were originally proposed by von Bertalanffy⁴, who suggested that growth rate is the difference between anabolic rate, B_a (biomass production, scaling as $m^{2/3}$), and catabolic rate, B_c (biomass breakdown, scaling as m): $dm/dt \propto B_a - B_c$. This case, which is considered by Banavar *et al.*, corresponds to $\alpha = 2/3$, $\beta = 1$, with a , b (and M) being 'arbitrary' parameters determined by fitting growth data. No explanation or derivation is given for any of these parameters, so the authors' version of equation (1) is simply a curve-fitting statistical description.

The assertion that $\alpha = 2/3$ by Banavar *et al.* is at odds with their earlier theoretical argument² for a $3/4$ exponent for B . In any case, von Bertalanffy's explanation (and that of Banavar *et al.*) for the origin of equation (1) cannot be correct, as both B_a and B_c scale as $m^{3/4}$, leading to $dm/dt \propto m^{3/4}$, or $m \propto t^4$ for all times.

To reveal the universality of growth that is implied by equation (1), we showed that, by plotting $r \equiv (m/M)^{1/4}$ against $\tau \equiv at/M^{1/4} - \ln[1 - (m_0/M)^{1/4}]$, all organisms conform to a predicted universal curve, $1 - e^{-\tau}$. Banavar *et al.* observe that a similar plot can be generated by using an unrealistic $\alpha = 2/3$, rather than $\alpha = 3/4$. Most data on ontogenetic growth are not of sufficient quality to distinguish between the two: we recognized this and made no claim that $\alpha = 3/4$ is a better fit than $\alpha = 2/3$. However, the statement by Banavar *et al.* that this curve is independent of α is misleading because r and τ depend explicitly on α , so the scaling curve cannot be constructed without knowing its value, as well as the values of a and b . (Indeed, Banavar *et al.* use our values based on a $3/4$ power.)

Banavar and colleagues' comment misses our central point that, because equation (1) is derived from fundamental principles concerning how growth is fuelled by metabolic power at the cellular level, many important quantities can be understood quantitatively. For example, our model elegantly interprets r as the proportion of total lifetime metabolic energy that is devoted to maintenance and other activities.

We further contend that the implication made by Banavar *et al.* that their equation $dm/dt = m^a f(m/M)$ is the most general form of the growth equation is also misleading. The function f depends on several variables, including m , M , B , m_c , E_c , B_c , cell growth and lifetime, time to maturity, and so on. Without a specific mechanistic model, why should f depend only on m/M , and what sets the fundamental timescale for growth? Our equation (1) answers these and other questions. It contains, derives and predicts many fundamental biological and physical variables that capture the essential features

of ontogenetic growth, yet it yields an extraordinarily simple universal equation.

Geoffrey B. West*, Brian J. Enquist†, James H. Brown†§

*Los Alamos National Laboratory, Los Alamos, New Mexico 87545, USA

e-mail: gbw@lanl.gov

†Santa Fe Institute, Santa Fe, New Mexico 87501, USA

‡Department of Ecology and Evolutionary Biology, University of Arizona, Tucson, Arizona 85721, USA

§Department of Biology, University of New Mexico, Albuquerque, New Mexico 87131, USA

1. West, G. B., Brown, J. H. & Enquist, B. J. *Nature* **413**, 628–631 (2001).
2. Banavar, J. R., Maritan, A. & Rinaldo, A. *Nature* **399**, 130–132 (1999).
3. West, G. B., Brown, J. H. & Enquist, B. J. *Science* **276**, 122–126 (1997).
4. von Bertalanffy, L. Q. *Rev. Biol.* **32**, 217–231 (1957).

COMMUNICATIONS ARISING

Climate change

Regional warming and malaria resurgence

Disease outbreaks are known to be often influenced by local weather, but how changes in disease trends might be affected by long-term global warming is more difficult to establish. In a study of malaria in the African highlands, Hay *et al.*¹ found no significant change in long-term climate at four locations where malaria incidence has been increasing since 1976. We contend, however, that their conclusions are likely to be flawed by their inappropriate use of a global climate data set. Moreover, the absence of a historical climate signal allows no inference to be drawn about the impact of future climate change on malaria in the region.

The findings of Hay *et al.*¹ are based on interpolation to four locations, from a 0.5° -resolution gridded climate data set^{2,3}, within an area of large altitudinal contrasts

and sparse, geographically dispersed historical climate data. In such a region, these gridded climate anomalies are often based on data outside a particular grid cell, and their interpolation to specific sites ignores local elevational dependencies. The mean site altitudes used by Hay *et al.* (1,693 m, 1,819 m, 1,893 m and 2,031 m), compared with those of the actual input weather-station sites (506 m, 1,110 m, 1,312 m, 1,515 m, 1,624 m and 1,635 m), differ on average by 575 m, which corresponds to a temperature deviation of 3°C . The sparse weather stations also range over a wide expanse of 5° latitude and 7° longitude. This climate data set is appropriate for up-scaling to African regions, but not for down-scaling to specific area locations; it cannot therefore support the type of analysis carried out by Hay *et al.*

Hay *et al.* focus on climate trends, but 'climate change' also applies to changes in variability. In regression analysis, a trend in covariates is not necessary; a change in variance can yield larger or more frequent responses. In the African highland, increases in the magnitude or frequency of malaria epidemics are most closely associated with short-term climate anomalies^{4–6}. Because of the existence of critical climate thresholds, the association between change in malaria incidence and change in climate can be biologically meaningful, even without 'significant' climate change.

Based on an understanding of the limitations of the gridded climate data set³ and on an examination of individual station data^{7–9}, we calculate that in the east African region encompassing the four study sites there was a mean warming trend of 0.15°C per decade during 1970–98, aggregated across the 320 0.5° grid boxes (Fig. 1). This regional warming tells us little, however, about climate trends at specific sites⁹, as these data^{2,3} are not designed to reveal such information.

In contrast to Hay *et al.*, we have identified regional warming trends in east Africa that parallel rising trends in malaria inci-

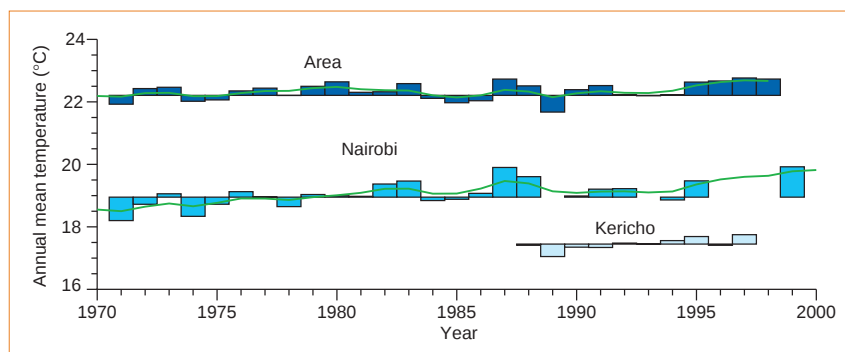


Figure 1 Annual mean temperature for Nairobi airport (WMO 63741, 1.3°S , 36.9°E , 1,624 m) and Kericho (0.37°S , 35.35°E , 2,031 m) are plotted as bars to show deviations from the averages for 1961–90 (19.0°C) and 1988–97 (17.4°C), respectively. These station records are complemented with a large-area average from a 0.5° -gridded time-series defined by 4°S , 4°N , 28°E , 38°E (refs 2,3). The area-averaged annual mean temperature is plotted (top) as bars that display deviations from the 1961–90 average (22.2°C). Time series for Nairobi airport and area-averaged data are also plotted after smoothing with a 10-year gaussian filter to emphasize changes on decadal timescales. In addition to the observed temperature trends, note the marked altitudinal dependency in temperatures.

dence at specific sites. But such climate and malaria data sets must be considered at comparable spatial and temporal scales. For example, in a comparison of monthly climate and malaria data in highland Kakamega, Kenya¹⁰, we found a close association between malaria transmission and monthly maximum temperature anomalies from 1997 to 2000, using data from the same location and over the same period of time. Hay and colleagues simply compared point-incidence rates with downscaled gridded climate data, rather than coincident longitudinal malaria and climate data.

We conclude that a reliable assessment of long-term relationships between climate and malaria incidence requires increased local monitoring of appropriate climate and disease variables to establish data sets that can support long-term trend analysis. Interdisciplinary teams are needed to analyse processes as diverse as climate and human disease.

Jonathan A. Patz*, **Mike Hulme†**, **Cynthia Rosenzweig‡**, **Timothy D. Mitchell†**, **Richard A. Goldberg‡**, **Andrew K. Githeko§**, **Subhash Lele||**, **Anthony J. McMichael¶**, **David Le Sueur#**

*Department of Environmental Health Sciences, Johns Hopkins University, Bloomberg School of Public Health, Baltimore, Maryland 21205, USA
e-mail: jpatz@jhsph.edu

†Tyndall Centre for Climate Change Research, School of Environmental Sciences, University of East Anglia, Norwich NR4 7TJ, UK

‡NASA, Goddard Institute for Space Studies, and Columbia University, New York, New York 10025, USA

§Climate and Human Health Research Unit, Centre for Vector Biology and Control Research, Kenya Medical Research Institute, PO Box 1578, Kisumu, Kenya

||Department of Mathematical and Statistical Sciences, University of Alberta, Edmonton AB T6G 2G1, Canada

¶National Centre for Epidemiology and Population Health, Australian National University, Canberra, ACT 0200, Australia

#South African Medical Research Council, PO Box 17120, Congella, Durban 4013, South Africa

elsewhere². Regarding the climate surfaces we used, the data set has superior spatial resolution compared with other data sets of similar temporal extent^{3,4}, and has been used to quantify climate change across Africa at $0.5 \times 0.5^\circ$ spatial resolution⁵ (although the resulting maps are smoothed to emphasize regional changes). It is inconsistent to assert that these same data are insufficient to demonstrate a lack of climate change.

Furthermore, these smoothed patterns⁵ are at odds with the marked and variable trends in temperature identified across east Africa using long-term temperature records from meteorological stations⁶. Events occurring at the $0.5 \times 0.5^\circ$ resolution will be less variable when averaged over wider areas, but we know of no evidence that climate surfaces interpolated from meteorological stations consistently fail to reveal trends in climate experienced at those locations.

Crucially, further work has confirmed a very high degree of correspondence between the climate surfaces^{3,4} and meteorological-station data from Kericho. Moreover, these station data show no significant trend in temperature or rainfall during the 1966–95 period over which complete longitudinal hospital records show malaria incidence to have increased significantly⁷.

The malaria resurgences documented at these four sites are not “point-prevalence rates”, but estimates from longitudinal records of health facilities, whose catchment populations range over continuous highland areas that are similar in size to a pixel of the climate surface that we used^{3,4}.

The sparse coverage of meteorological stations in the data set^{3,4} before 1910 in the east African region is problematic⁴, and these data were excluded from our analyses. The full 1901–95 data set was used by one of the correspondents, however, in their trend analyses of African climate⁵. Moreover, our conclusions remain unaltered in the light of tests repeated for the 1970–95 period, which is coincident with the malaria resurgences and represents the most consistent meteorological-station coverage.

Patz *et al.* estimate a temperature deviation of 3°C arising from the difference in the mean elevation of the climate surface pixels^{3,4} and the elevation of meteorological station sites that contribute to them. However, the thin-plate spline procedure⁸ used to generate the 1961–90 climate normals³ that formed the basis of the entire climate data set⁴ explicitly takes account of altitude to correct for the elevational dependency of temperature.

Application of our methodology¹ to the regional time-series cited by Patz *et al.* shows that the purported warming trend is not significant (details available from the authors). It is critical to test for climate changes that coincide in both space and time with the disease data^{7,9}, so changes on regional scales or at distant sites (such as

Kakamega and Nairobi) are irrelevant, whether or not they achieve significance.

The more subtle impacts of non-significant long-term changes in climate on malaria incidence deserve to be investigated, but have not been demonstrated, so we cannot attribute significant increases in malaria incidence to non-significant changes in climate. It is because malaria incidence is related to climate that any study of long-term climate change must factor out seasonality, unlike the cited study on seasonal variability over four years¹⁰. Finally, windowed Fourier analysis of the meteorological station data for Kericho also showed no change in annual temperature or rainfall variability since 1966 (ref. 11), a conclusion that is corroborated by global-scale analysis of climate variability during this century¹².

Rather than climate change, variations in environmental, social and epidemiological factors are more plausible explanations for the malaria resurgences at the four sites we examined¹ and at three others in Ethiopia, Madagascar and Tanzania¹³. Evidence against the epidemiological significance of climate change in the recent malaria resurgences in Africa is mounting^{7,9,13} and remains unmatched by any contrary evidence.

Simon I. Hay*, **Jonathan Cox‡**, **David J. Rogers***, **Sarah E. Randolph§**, **David I. Stern||**, **G. Dennis Shanks¶**, **Monica F. Myers#**, **Robert W. Snow†****

*TALA Research Group, and §Oxford Tick Research Group, Department of Zoology, University of Oxford, Oxford, OX1 3PS, UK

e-mail: simon.hay@zoo.ox.ac.uk

†Kenya Medical Research Institute/Wellcome Trust Collaborative Programme, PO Box 43640, Nairobi, Kenya

‡Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London WC1E 7HT, UK

||Department of Economics, Rensselaer Polytechnic Institute, Troy, New York 12810, USA

¶US Army Medical Research Unit, PO Box 30137, Nairobi, Kenya

#Decision Systems Technologies, Rockville, Maryland 200850, USA

**Centre for Tropical Medicine, University of Oxford, John Radcliffe Hospital, Oxford OX3 9DU, UK

1. Hay, S. *et al.* *Nature* **415**, 905–909 (2002).

2. New, M., Hulme, M. & Jones, P. *J. Clim.* **12**, 829–857 (1999).

3. New, M., Hulme, M. & Jones, P. *J. Clim.* **13**, 2217–2238 (2000).

4. Gilles, H. in *Bruce Chwatt's Essential Malaria* (eds Gilles, H. & Warrell, D.) 124–163 (Arnold, London, 1993).

5. Patz, J. *et al.* *J. Trop. Med. Int. Health* **3**, 818–827 (1998).

6. Cox, J., Craig, M., Le Sueur, D. & Sharp, B. *Mapping Malaria Risk in the Highlands of Africa* (MARA, London School of Hygiene and Tropical Medicine, London, 1999).

7. Climate Library, March 2002 (accessed);

www.giss.nasa.gov/data/update/gistemp

8. Peterson, T. & Vose, R. *Bull. Am. Meteorol. Soc.* (1997).

9. Pielke, R. A. *Int. J. Climatol.* **22**, 421–434 (2002).

10. Githeko, A. & Ndegwa, W. *Global Change Hum. Health* **2**, 54–63 (2001).

Hay *et al.* reply — In our study¹ we did not address the impact of predicted climate change on malaria incidence because a complete global assessment is reported

1. Hay, S. I. *et al.* *Nature* **415**, 905–909 (2002).

2. Rogers, D. J. & Randolph, S. E. *Science* **289**, 1763–1766 (2000).

3. New, M., Hulme, M. & Jones, P. *J. Clim.* **12**, 829–857 (1999).

4. New, M., Hulme, M. & Jones, P. *J. Clim.* **13**, 2217–2238 (2000).

5. Hulme, M., Doherty, R., Ngara, T., New, M. & Lister, D. *Clim. Res.* **17**, 145–168 (2001).

6. King'uyu, S. M., Ogallo, L. A. & Anyamba, E. K. *J. Clim.* **13**, 2876–2886 (2000).

7. Shanks, G. D., Hay, S. I., Stern, D. I., Biomndo, K. & Snow, R. W. *Emerging Infect. Dis.* **8**, 1404–1408 (2002).

8. Hutchinson, M. F. *Int. J. Geogr. Inf. Syst.* **9**, 385–403 (1995).

9. Hay, S. I. *et al.* *Direct. Sci.* **1**, 82–85 (2002).

10. Githeko, A. & Ndegwa, W. *Global Change Hum. Health* **2**, 54–63 (2001).

11. Rogers, D. J., Randolph, S. E., Snow, R. W. & Hay, S. I. *Nature* **415**, 710–715 (2002).

12. Vinnikov, K. Y. & Robock, A. *Geophys. Res. Lett.* **29**, 10.1029/2001GL014025 (2002).

13. Hay, S. I. *et al.* *Trends Parasitol.* (in the press).

Rho GTPases in cell biology

Sandrine Etienne-Manneville* & Alan Hall*†

* MRC Laboratory for Molecular Cell Biology and Cell Biology Unit, Cancer Research UK Oncogene and Signal Transduction Group; and † Department of Biochemistry and Molecular Biology, University College London, Gower Street, London WC1E 6BT, UK

Rho GTPases are molecular switches that control a wide variety of signal transduction pathways in all eukaryotic cells. They are known principally for their pivotal role in regulating the actin cytoskeleton, but their ability to influence cell polarity, microtubule dynamics, membrane transport pathways and transcription factor activity is probably just as significant. Underlying this biological complexity is a simple biochemical idea, namely that by switching on a single GTPase, several distinct signalling pathways can be coordinately activated. With spatial and temporal activation of multiple switches factored in, it is not surprising to find Rho GTPases having such a prominent role in eukaryotic cell biology.

GTPases are molecular switches that use a simple biochemical strategy to control complex cellular processes. They cycle between two conformational states: one bound to GTP ('active' state), the other bound to GDP ('inactive' state), and they hydrolyse GTP to GDP (Fig. 1). In the 'on' (GTP) state, GTPases recognize target proteins and generate a response until GTP hydrolysis returns the switch to the 'off' state. This idea has been elaborated throughout evolution, and mammalian cells contain several hundred GTPase switches. The Ras superfamily of GTPases is particularly interesting to cell biologists, as its members have turned out to be master regulators of many aspects of cell behaviour. These small, monomeric GTPases, which number over 60 in mammals, fall into five major groups: Ras, Rho, Rab, Arf and Ran. This review focuses specifically on Rho GTPases, emphasizing the insights they are providing into the molecular mechanisms underlying cell biology.

Although the Rho switch itself is straightforward, it is very carefully regulated, and the human genome contains over 60 activators (guanine nucleotide exchange factors, GEFs) and over 70 inactivators (GTPase-activating proteins, GAPs) for this family (Fig. 1). Unfortunately, target proteins do not contain a single recognizable sequence motif useful in database searches, but for Rho, Rac and Cdc42, the three best-characterized members of the family, over 60 targets have so far been identified experimentally. This is unparalleled complexity for a GTPase switch and points to their importance in the eukaryotic cell.

The *Rho* gene was identified in 1985, but it was observations reported in 1992 that provided the first insights into the cellular function of Rho GTPases. Constitutively activated (GTPase deficient) mutants of Rho and Rac were found to induce the assembly of contractile actin and myosin filaments (stress fibres) and actin-rich surface protrusions (lamellipodia), respectively, when introduced into fibroblasts^{1,2}. Later, Cdc42 was shown to promote the formation of actin-rich, finger-like membrane extensions (filopodia)^{3,4}. The conclusion that Rho, Rac and Cdc42 regulate three separate signal transduction pathways linking plasma membrane receptors to the assembly of distinct filamentous actin structures has since been confirmed in a wide variety of mammalian cell types as well as in yeast, flies and worms. Few functional data are currently available on the other 13 members of the mammalian Rho GTPase family.

Further interest in Rho GTPases has been fuelled by the realization that they regulate many other signal transduction pathways in addition to those linked to the actin cytoskeleton. They participate in the regulation of cell polarity, gene transcription, G1 cell cycle progression, microtubule dynamics, vesicular transport pathways and a variety of enzymatic activities ranging from an NADPH oxidase in phagocytes to a glucan synthase in yeast. The seemingly endless list of activities is confusing, but is consistent with the large

number of target proteins identified. Our aim is to integrate this information into a cell biological context, and we have organized the review into three broad, often overlapping topics relevant to all eukaryotic cells: morphology, movement and behaviour.

Morphology

Animal cells adopt a huge diversity of shapes, ranging from the relatively simple-looking columnar epithelial cell through to the highly complex branched structure of a neuron. Shape is highly dependent on the external environment, which also provides directional cues that drive the establishment of intracellular polarity.

Establishing intracellular asymmetry

Even unicellular organisms respond to directional cues, and genetic analysis of budding yeast provided the first evidence linking Cdc42 to cell polarity⁵. In the absence of this GTPase, *Saccharomyces cerevisiae* cannot establish a defined site for daughter cell growth and, as a consequence, cells expand isotropically (Fig. 2). Similarly, during mating, yeast cells extend surface protrusions (shmoo) directed along a gradient of secreted pheromone; without Cdc42p (yeast Cdc42) the protrusions are no longer oriented correctly. The conclusion is clear: Cdc42p is not needed for cell growth or protrusion formation *per se*, but it is needed for these processes to

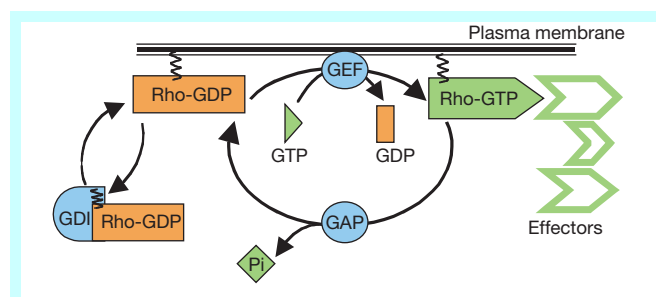


Figure 1 The Rho GTPase cycle. Twenty mammalian Rho GTPases have been described: Rho (three isoforms: A, B, C); Rac (1, 2, 3); Cdc42; TC10; TCL; Chp (1, 2); RhoG; Rnd (1, 2, 3); RhoBTB (1, 2); RhoD; Rif; and TTF. They cycle between an active (GTP-bound) and an inactive (GDP-bound) conformation. In the active state, they interact with one of over 60 target proteins (effectors). The cycle is highly regulated by three classes of protein: in mammalian cells, around 60 guanine nucleotide exchange factors (GEFs) catalyse nucleotide exchange and mediate activation; more than 70 GTPase-activating proteins (GAPs) stimulate GTP hydrolysis, leading to inactivation; and four guanine nucleotide exchange inhibitors (GDIs) extract the inactive GTPase from membranes. All Rho GTPases are prenylated at their C terminus, and this is required for function. Rnd proteins are exceptional in that they do not hydrolyse GTP *in vitro*, which is an unusual property for a regulatory GTPase. It has been argued that they are controlled by expression, but it is equally possible that a yet to be identified GAP is required for hydrolysis.

occur at the right place on the cell surface. Once the bud site has been established, the organization of the actin and microtubule cytoskeletons and the direction of vesicular transport pathways are fixed (Fig. 2). In a subsequent step, Rho1p (another member of the yeast Rho GTPase family) is recruited and activates three signal transduction pathways required for growth of the bud⁶.

These observations provide a number of potential models for morphogenesis in more complicated systems. First, Cdc42 is the principal determinant in establishing correct cell polarity with respect to the external environment. Second, establishing intracellular asymmetry is a prerequisite for morphogenesis. Third, small GTPases provide distinct, but cooperative contributions to the morphogenetic process. The biochemistry of polarity establishment has proved more difficult to elucidate, and one important distinction to bear in mind is that generating cellular asymmetry is not necessarily the same as generating asymmetry in the correct orientation. Thus, in yeast lacking Cdc42p, pheromone induces an asymmetrical, stochastic accumulation of shmoo components on the membrane surface, but these are transient. The role of Cdc42 is to stabilize this activity in the correct location with respect to the external cue. Positive feedback loops probably have a prominent role, and in agreement with this, a target of Cdc42p (Bem1p) that localizes to the shmoo stabilizes the GEF (Cdc24p) that is responsible for activating Cdc42p in the first place⁷.

The genetic analysis of the first cell (zygote) division during *Caenorhabditis elegans* development has provided some of the best insights into how cellular asymmetry can be generated. The protein products of six *par* genes (PAR-1–6, partitioning defective) and an atypical protein kinase C (PKC-3) are essential for establishing the anterior/posterior axis within this single cell. PAR-1 and PAR-2

localize at the posterior end of the cell (where fertilization takes place), whereas PAR-3, PAR-6 and PKC-3 localize at the anterior end (Fig. 2)⁸. PAR proteins are conserved throughout the animal kingdom (although not in yeast) and the PAR-6–atypical PKC complex has been shown to direct cell asymmetry during oocyte maturation and neuroblast division in *Drosophila* (Fig. 2)⁹. Further work in *C. elegans* has revealed that inhibition of Cdc42 completely disrupts polarity and delocalizes all PAR proteins^{10,11}. A clue to its role comes from mammalian cells, where Cdc42GTP interacts directly with Par6 (mammalian homologue of PAR-6) to stimulate the kinase activity of the associated atypical PKC^{12–14}. It is likely that a spatially localized, active Cdc42GTP–PAR-6–PKC-3 complex is required for maintaining the asymmetry of the zygote in *C. elegans* (Fig. 2). These examples relate to the establishment of cell polarity before cell division and are required for cell fate determination, but are Cdc42 and Par proteins used to establish polarity during morphogenesis?

Establishing shape

Typical epithelial cells, such as those lining the colon or those forming the *Drosophila* wing, form monolayers of packed cuboidal cells with specialized cell–cell contacts and a distinctive asymmetrical distribution of proteins at the basolateral and apical membranes (Fig. 2). The establishment of epithelial cell morphology is driven primarily by two types of intercellular adhesive junction: adherens junctions, which form a strong mechanical link between adjacent cells, and tight junctions, which form a physical barrier preventing the diffusion of both lipids and proteins between the basolateral and apical membranes. Accompanying the assembly of these junctions, the actin and microtubule cytoskeletons are re-organized and vesicular transport is polarized, and in this way, the

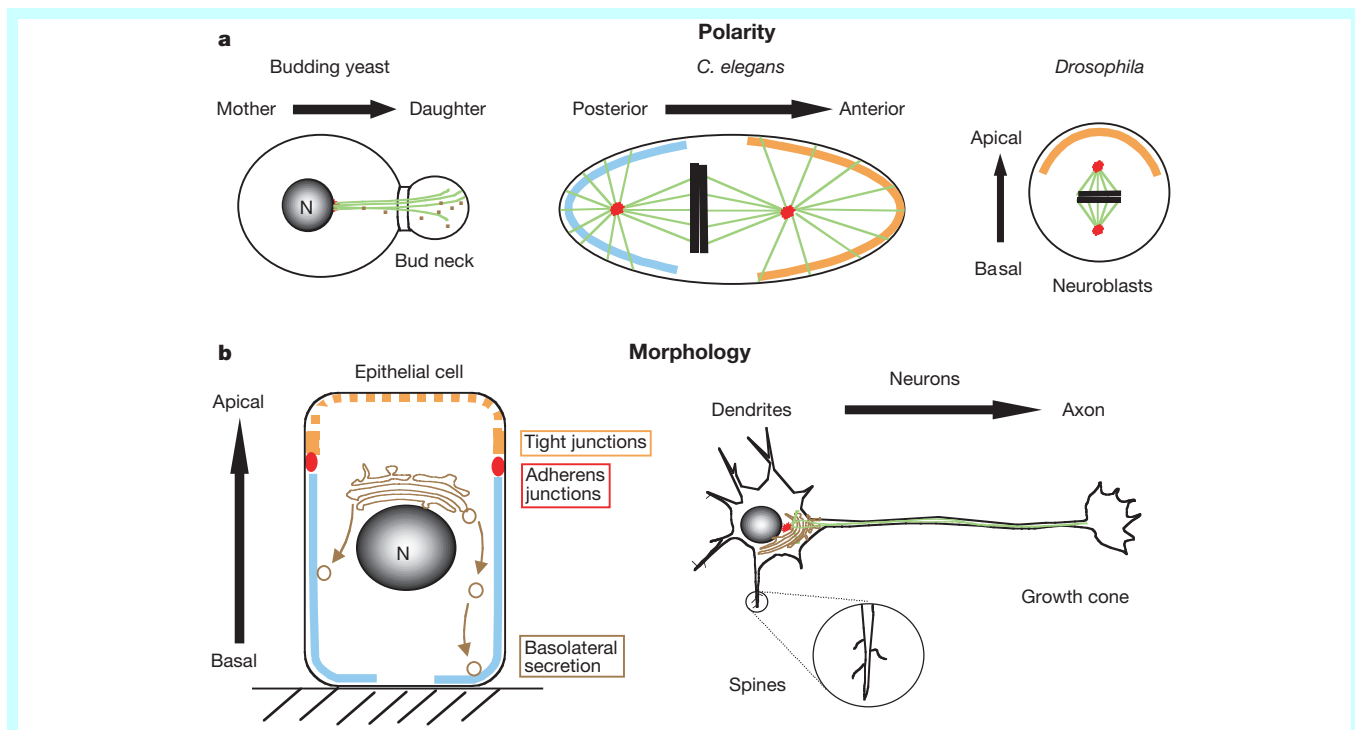


Figure 2 Morphogenesis. **a**, Polarity. All eukaryotic cells are able to polarize in response to cues at the plasma membrane. Cdc42 is required for bud site assembly in yeast and, if deleted, cells expand isotropically. The polarized distribution of proteins in the *C. elegans* zygote or in *Drosophila* neuroblasts leads to asymmetrical cell division, a prerequisite for cell fate determination. In *C. elegans*, polarity establishment is controlled by Cdc42, the products of six *par* genes (PAR-1–6), and an atypical protein kinase C (PKC-3). PAR-6–PAR-3–PKC-3 localize at the anterior pole (orange line); PAR-1 and PAR-2 localize at the posterior pole (blue line). The microtubule organizing centre (MTOC) is shown in red and microtubules are in green. **b**, Morphology. Polarity is

also a major factor in determining cell morphology. In an epithelial cell, Cdc42, Par6, Par3 and atypical PKC regulate tight junction assembly to determine apical/basolateral polarity. Cdc42, Rac and Rho are also involved in the assembly of adherens junctions and, along with basolaterally directed vesicle transport, this leads to the typical columnar shape. Junctions are further stabilized by associated actin filaments and, here too, Rho GTPases have an involvement. Rho GTPases are involved in determining the morphology of neurons; they affect axon, dendrite and spine growth, as well as axon guidance. It has not yet been shown whether Cdc42 has a role in establishing the polarity of the neuron.

shape of the epithelial cell is established and maintained.

During the formation of epithelia, cadherin-mediated cell–cell contact first leads to the development of adherens junctions. How this occurs is beginning to be unravelled¹⁵. First, Cdc42 and Rac are activated by and recruited to cadherin–cadherin contact sites formed between neighbouring cells. Observations made with epidermal cells during *C. elegans* development, ectodermal cells during *Drosophila* dorsal closure, and cultured mouse keratinocytes have revealed that filopodia and/or lamellipodia emerge from cells and penetrate into neighbouring cells^{16–18}. This seems to provide a driving force to generate intimate cell–cell contact, and Cdc42 and Rac probably have a prominent role in this context. Inhibition of Rho also prevents adherens junction assembly; its contribution may be to stabilize junctions through actin filament assembly; however, non-actin-dependent pathways have also been proposed.

The formation of the tight junction is required to establish the apical/basolateral asymmetry (Fig. 2). Notably, a Par6–atypical PKC complex is required for the formation of this junctional structure and Par3 is recruited to initial cell–cell contacts and remains associated with tight junctions in polarized cells¹⁹. The activity of this complex is probably regulated by Cdc42, although experimental evidence to support this is contradictory^{20,21}. These observations point to a conserved mechanism for establishing intracellular asymmetry in both the *C. elegans* zygote and the mammalian epithelial cell. Although the precise role of Cdc42 in tight junction assembly is not clear, one possibility is that activation of atypical

PKC induces phosphorylation of Par3, allowing this scaffold-like protein to interact with junctional components, such as Jam²². Unlike Cdc42, Par6 and atypical PKC are not required for adherens junction assembly. Finally, epithelial cell polarity also depends on the cell's interaction with extracellular matrix through integrins at the basal surface (Fig. 2). The formation of hollow cysts (apical-in, basolateral-out) when MDCK cells are placed in a three-dimensional matrix, for example, requires recognition of laminin. Interestingly, Rac is required for the correct assembly of extracellular laminin and therefore influences the orientation of the apico-basal axis²³. This points to an intimate cooperation between Rac and Cdc42 in establishing epithelial cell polarity.

Rho GTPases contribute to epithelial cell morphogenesis through other activities. Disruption of Cdc42 in MDCK monolayers results in the inhibition of vesicular transport specifically to basolateral surfaces, leading to a selective depolarization of basolateral membrane proteins²⁴. Interestingly, the delivery of vesicles containing basolateral membrane proteins is mediated by the exocyst complex, and in yeast, Cdc42p interacts directly with one of its components, Sec3p²⁵. In mammalian cells, Ral, a small GTPase in the Ras family, interacts with another exocyst component, Sec5, and Ral has been reported to be activated by Cdc42 (refs 26, 27). This could be a mechanism for delivering membrane to the lateral edges, leading to cell elongation and the formation of the typical columnar shape.

The generation of intracellular asymmetry in response to environmental cues is required in most animal cell types to direct

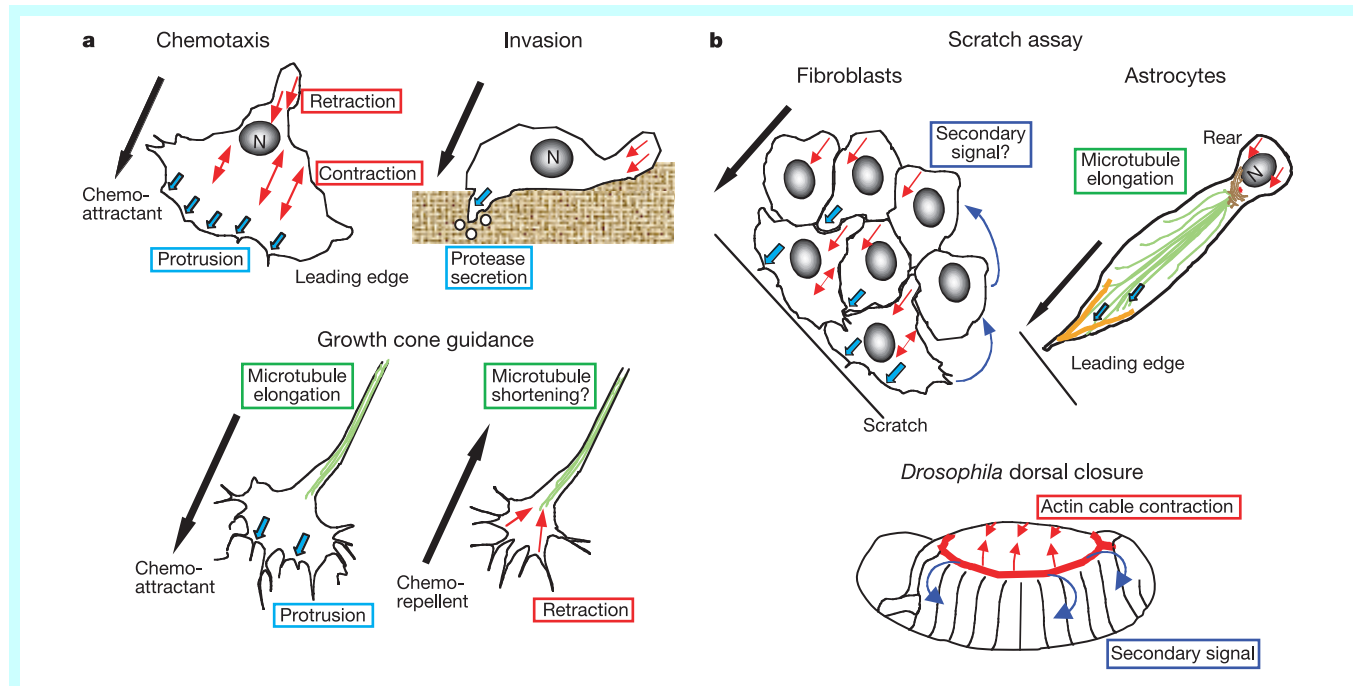


Figure 3 Movement. **a**, Independent movement. Cells move through the polarized and dynamic re-organization of the actin cytoskeleton, involving a protruding force at the front (blue arrows), combined with a contractile force in the cell body (double-headed red arrows). This contractile activity leads to retraction of the rear of the cell as the adhesions are lost (single-headed red arrows). Rho GTPases act spatially and temporally to control all these aspects. Rac regulates actin polymerization at the front to promote protrusion. Cdc42 acts at the front to control direction in response to extracellular cues. Rho stimulates actin-myosin contraction in the cell body. Invading cancer cells are probably not directed by outside signals, but deregulated Rac is thought to have an important role. Invasion probably involves changes in gene expression, secretion of proteases, alterations in surface adhesion proteins, and loss of adhesion-dependent signals controlling cell death. Growth cone movement of an axon is governed by attractive or repulsive cues, leading to actin-dependent protrusion or retraction, respectively. Microtubule dynamics and vesicular transport pathways also have a role in facilitating directed cell migration. **b**, Coordinated movement. In *in vitro* scratch assays,

cells sense the free space left by the scratch and migrate together as a sheet. Rac is essential for forward movement in fibroblasts and astrocytes. Migrating cells are oriented perpendicularly to the scratch, most noticeably with astrocytes, showing pronounced elongation, with their MTOC (red circle) and Golgi apparatus (brown stack) in front of the nucleus. These aspects of cell polarity are controlled by Cdc42. In astrocytes, Par6 and PKC ζ are localized at the leading edge of migrating cells (orange line) and are essential for polarity (although not elongation). The movement of astrocytes is dependent on microtubule as well as actin dynamics. How cells behind the front row migrate coordinately with front row cells is unclear, but may involve secretion of soluble factors and/or mechanical tension. In *Drosophila* dorsal closure, ectodermal cells move coordinately as a sheet over the embryo. The Rho-dependent contraction of an actin ring (red line), joining all the cells of the front row, is involved. The Rac/Cdc42-dependent secretion of a soluble TGF β -like factor (Dpp, in blue) is required to promote migration of the entire monolayer. When cells approach each other, Cdc42- and Rac-dependent filopodia and lamellipodia protrude to execute closure.

a morphogenetic programme, but perhaps none is as complex as that in neurons. Neurons extend neurites, one of which differentiates into an axon, whereas the others become dendrites (Fig. 2). Dendrites form spines: small actin-rich projections along their length that form synapses with axons of other neurons. The mechanism that underlies the decision to form one axon from many neurites is not understood, but this is the step that establishes the polarity and morphology of the neuron. *In vitro* this seems to be a stochastic decision, but *in vivo* the developing neuron probably takes cues from its environment. Cdc42 and Par gene products may be involved in establishing the polarity of the differentiating neuron, but this has not yet been addressed.

Cdc42 and Rac are positive regulators of neurite outgrowth, whereas Rho inhibits neurite extension²⁸. On the basis of work in fibroblasts, where Cdc42 and Rac promote membrane protrusion (through actin filament assembly at the periphery) whereas Rho promotes membrane retraction (through contractile actin and myosin filaments), the opposing activities of these GTPases on the actin cytoskeleton have been used extensively to explain their observed effects in neurons. Thus, Rho is required to limit dendritic outgrowth in *Drosophila*²⁹, whereas Rac is required for nerve growth factor (NGF)-induced axon growth³⁰. Similarly, Rac activation increases the number of dendritic spines, whereas Rho activation has the opposite effect^{31,32}. The idea that opposing Rac and Rho activities control neuronal shape through the actin cytoskeleton may, however, turn out to be an oversimplification. It is still unclear, for example, what the relative contributions are of the actin and microtubule cytoskeletons in dendrite and axon outgrowth, and

there is now overwhelming evidence that Rho GTPases directly control microtubule dynamics (see below).

Movement

The migration of cells, either as individuals or as groups, is a principal feature of metazoan embryonic development, and in the adult it is required to maintain tissue integrity (for example, the movement of differentiating stem cells in the colon) and during immune surveillance (for example, neutrophil chemotaxis).

Single-cell migration

The ability of a cell to move requires the asymmetrical organization of cellular activities. The front of the migrating cell generates protrusive force, generally associated with the extension of a lamellipodium in the direction of migration coupled with the development of new cell adhesions to the extracellular substrate. This in itself is not sufficient, however, and cell contractility is required to allow the body and rear of the cell to follow the extending front³³. Through its ability to promote both protrusion and cell contraction, the actin cytoskeleton is believed to provide the driving force for cell migration (Fig. 3a).

Rac induces actin polymerization and integrin adhesion complex assembly at the cell periphery, leading to membrane protrusion, and it is essential for the migration of all cells examined so far³⁴. The biochemical mechanisms by which Rac catalyses actin polymerization are a focus of great interest. Four Rac targets (IRSp53, phosphatidylinositol-4-phosphate 5-kinase, p65Pak and LIM kinase) have been implicated, as has the Arp2/3 complex, which

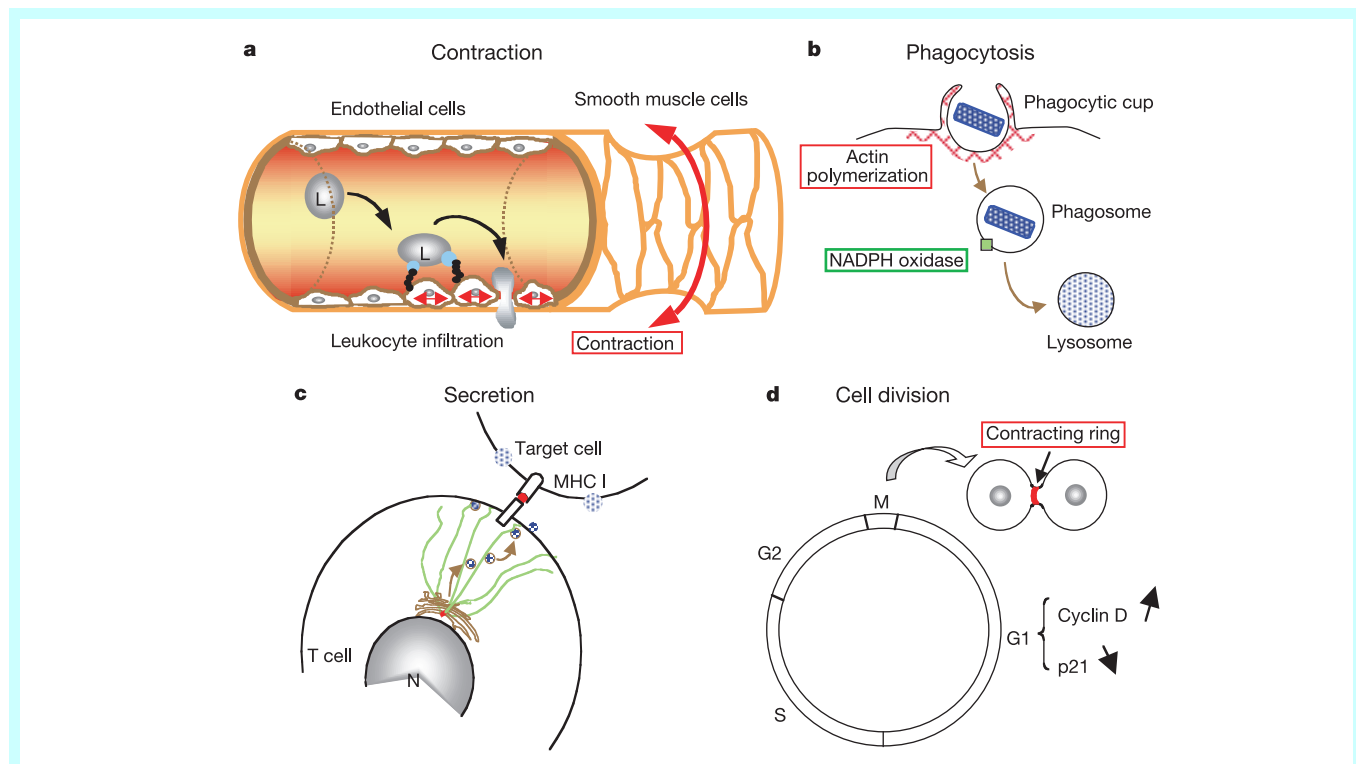


Figure 4 Behaviour. **a**, Contraction. Rho and its target p160Rho kinase are principal mediators of cell shape changes through their effects on actin and myosin contraction. Rho stimulates smooth muscle cell contraction (to regulate blood circulation) and endothelial cell contraction (to allow movement of blood cells into the surrounding tissues; extravasation). **b**, Phagocytosis. The localized polymerization of actin at peripheral sites can promote internalization of attached particles and microorganisms (blue rectangle). During type I phagocytosis, Cdc42 and Rac induce actin polymerization to form membrane extensions that engulf the particle. Rac also associates with p67phox, a component of the NADPH oxidase enzyme complex (green square), to stimulate the production of superoxide ions as part of a bactericidal

programme. **c**, Secretion. The re-orientation of the microtubule cytoskeleton (tubulin, green; MTOC, red) in cytotoxic T cells allows the targeted delivery of perforin-containing granules (blue dots) to the target APC. Cdc42 is essential for establishing cell polarity in this context, but its downstream targets have not been identified. **d**, Cell division. Rho, Rac and Cdc42 contribute different activities to the G1 phase of the cell cycle. Rho is required to prevent expression of p21, an inhibitor of G1 cyclin/Cdk, whereas Rac and Rho promote transcription and translation of cyclin D. Rho and Cdc42 are also required late in the cell cycle for the correct assembly of the actin and myosin contractile ring (red) that separates daughter cells.

catalyses *de novo* nucleation of actin polymerization and the formation of new filament branches³⁵. How Rac promotes the small integrin adhesion complexes found at the front of a migrating cell that are required to generate traction forces is unknown.

The protrusive activity required for cell migration must be spatially controlled to promote net cell translocation. FRET (fluorescence resonance energy transfer) microscopy has revealed that RacGTP levels are highest at the leading edge of a migrating cell. The mechanism of localized Rac activation is not clear, although integrin–matrix interactions probably play an important part³⁶. Work in *Drosophila*^{37,38} and *C. elegans*³⁹ has revealed the importance of a Dock 180–CrkII scaffold complex in mediating Rac activation by chemotactic agents during border cell and distal tip cell migration, respectively. The mammalian orthologues of these associate with integrins⁴⁰. The spatial distribution of Rho activity during cell migration has yet to be visualized, although in motile monocytes, it is responsible for contraction and retraction within the trailing cell body, suggesting that RhoGTP may be restricted to the cell body and excluded from the leading edge⁴¹.

Cell migration is normally directed and controlled by extracellular cues (Fig. 3). In tissue culture, many cells can adopt a polarized morphology with a front and rear, and are able to migrate, but this is only transiently sustained, leading essentially to a random walk (chemokinesis). The stabilization of directional movement (chemotaxis) requires input from external cues, and this is controlled by Cdc42. The best study of this so far is in macrophage cells moving up a gradient of a chemotactic factor. When Cdc42 is inhibited, the macrophage reverts to a random walk, whereas inhibition of Rac blocks all cell movement⁴². It seems that Cdc42 directs and/or stabilizes Rac activity at the cell front. How it does this is not known, but it will be interesting to see whether the Par proteins participate as they do in migrating sheets of cells (see below).

Although localized Rac-induced actin polymerization is the driving force, migration may be facilitated by other cellular activities. The microtubule cytoskeleton, for example, is highly polarized during migration. This is characterized by the accumulation of stabilized, detyrosinated microtubules in the vicinity of the leading edge; increased microtubule growth at the leading edge; and re-orientation of the microtubule organizing centre (MTOC) either in front of or behind the nucleus, with respect to the direction of migration⁴³. Notably, Rho (acting through p160Rho kinase and mDia) promotes the accumulation of detyrosinated microtubules, Rac (acting through p65Pak) inactivates the microtubule destabilizing protein, stathmin and Cdc42 regulate the orientation of the MTOC.

Axon guidance can be thought of as a specialized form of directed migration: the cell body does not move, but the growth cone at the axon tip responds to attractive and repulsive extracellular cues as it extends to its target site (Fig. 3a)⁴⁴. Repulsion induced by the ephrin/Eph family of ligand/receptors requires Rho activation, whereas growth cone advance is associated with Rac activation. This fits well into the mutually opposing retraction/extension model previously mentioned for Rho and Rac. However, all is not as it seems; another class of chemorepellent, the semaphorins, acting through the plexin receptor family require Rac activity to induce growth cone collapse. The clearest *in vivo* work implicating Rho GTPases in the regulation of growth cone guidance has come from genetic screens in *Drosophila*⁴⁵ and *C. elegans*⁴⁶ showing that Rac is essential for axons to find their correct target site. If directional cues within the growth cone are at all similar to those operating in chemotactic cells, then it is to be expected that Cdc42 will also have a crucial role. So far, however, there is little direct evidence for this.

Coordinated cell migration

During embryogenesis, many cells move as sheets or as loosely associated groups and not as individuals. One of the best-studied *in vivo* examples of this is during *Drosophila* dorsal closure (Fig. 3b).

Here, cell migration and cell stretching are not so easily distinguishable, and the morphogenetic movements seem to have more similarity to embryonic wound healing where the formation of an actin and myosin cable, which joins all cells at the leading edge, promotes movement through a ‘purse-string’ contraction mechanism¹⁸. Here, Rho (acting through PKN and p160Rho kinase) controls the assembly and the contraction of the actin and myosin cable that drives wound or dorsal closure^{47,48}. Rac and Cdc42 are also required during dorsal closure and they contribute at least two activities. First, as the two migrating cell sheets come close, they promote filopodia and lamellipodia, which interact with opposing cells and facilitate completion of dorsal closure¹⁸. Second, they regulate the Jun N-terminal kinase (JNK) mitogen-activated protein (MAP) kinase cascade required for transcription of a transforming growth factor- β family member, Dpp, in cells at the leading edge⁴⁹. Secreted Dpp acts in a paracrine fashion to stimulate coordinated movement of the epithelial sheet⁵⁰.

One of the simplest ways of generating mammalian cell migration *in vitro*, by scratching (wounding) a cell monolayer, induces the coordinated movement of sheets of cells, at least in non-transformed cells (Fig. 3b). Using either primary fibroblasts or primary astrocytes, Rac is essential for cell migration in these assays. As with single cells, migrating sheets of cells recognize the direction of migration (the space formed by the scratch) and polarize so that protrusive activity is restricted to the front, the MTOC re-orientates in front of the nucleus to face the direction of migration, and secretion is directed towards the front. Cells in a monolayer or a group have an additional directional cue that is provided by interactions with their neighbours. Although the mechanisms operating within a cell to promote migration are becoming clearer, those promoting coordinated migration are poorly understood (Fig. 3b).

In both astrocyte and fibroblast scratch assays, inhibition of Cdc42 leads to misdirected protrusive activity and a random orientation of the MTOC^{51–53}. In astrocytes at least, both these aspects of polarity depend on a Par6–atypical PKC complex localized at the leading edge. The initiating external cue in this assay is integrin engagement (with matrix) at the front of leading edge cells, causing Cdc42 activation. Whether the positive, Cdc42-dependent cues found at the front of the migrating astrocytes are reinforced by ‘negative’ cues emanating from points of cell–cell contact with neighbours is not known. Thus, the molecules that control the establishment of polarity in migrating monolayers of astrocytes are the same as those that control asymmetrical cell division in *C. elegans* and tight junction assembly in epithelial cells.

Astrocytes undergo marked cell elongation (perpendicular to the scratch) during cell migration, which is also Rac-dependent (Fig. 3b). Unlike migration, however, this shape change is independent of actin polymerization, but is instead dependent on microtubule dynamics. It is unknown whether microtubule extension itself generates the elongated shape, or whether it is something (perhaps vesicular transport) that depends on microtubule dynamics. Rho has also been reported to induce a microtubule-dependent shape change (mediated by its target mDia) in non-migrating HeLa cells⁵⁴.

Behaviour

Mammalian Rho GTPases are well suited to control many of the specialized functions in cells of the adult. A brief overview will be presented here to demonstrate the diversity of effects uncovered so far, but more thorough reviews of this area can be found elsewhere^{15,55}.

Cell contraction

Contractile actin and myosin filaments are used by many cell types to induce rapid, reversible changes in shape. For example, vascular smooth muscle cells control blood flow by constricting and dilating in response to physiological stimuli (Fig. 4). Numerous agonists induce Rho activation in aortic smooth muscle cells, which through

p160Rho kinase stimulates myosin light chain phosphorylation to promote actin and myosin contraction^{56,57}. Small molecule inhibitors of p160Rho kinase prevent agonist-induced contraction of both vascular and bronchial smooth muscle cells, and have been reported to reduce high blood pressure in a rat model of hypertension⁵⁸. Similarly, Rho and p160Rho kinase control the barrier function of vascular endothelial cells to regulate extravasation of lymphocytes from circulating blood into surrounding tissues (Fig. 4). Addition of vascular endothelial growth factor (VEGF) or thrombin to endothelial cells in culture induces an increase in the permeability of cell monolayers and this is partially inhibited by p160Rho kinase inhibitors⁵⁹. It is thought that contraction forces, generated through increased actin and myosin activity, are responsible for destabilizing the endothelial cell–cell junctions⁶⁰.

Phagocytosis

The assembly and disassembly of peripheral actin filaments can be used to promote localized changes in the structure of the plasma membrane; phagocytosis is driven in this way (Fig. 4). Two distinct phagocytic pathways have been uncovered in the mammalian macrophage: type I (for example, through the immunoglobulin receptor) requires both Rac and Cdc42, whereas type II (for example, through the complement receptor) requires Rho⁶¹. In *C. elegans*, phagocytosis of apoptotic cells during development occurs through two distinct pathways, one of which requires Rac⁶². It is not clear why phagocytosis needs alternative pathways, but the answer probably lies not in the uptake process itself, but in other associated activities. For example, immunoglobulin, but not complement-mediated phagocytosis, is accompanied by activation of the NADPH enzyme complex, which coincidentally is allosterically regulated by Rac (Fig. 4)⁶³. The actin-rich surface protrusions induced by Rac and Cdc42 also mediate a related endocytic process, pinocytosis. Not only is this important for the uptake of essential nutrients, but it is used by immature dendritic cells to sample the surrounding tissue environment. Proteins taken up by Cdc42/Rac-dependent pinocytosis are processed and presented as peptides associated with major histocompatibility complex molecules, on the surface of mature dendritic cells⁶⁴.

The central role of Rho GTPases in phagocytosis provides at least a partial explanation of why they are such common substrates for bacterial toxins and effectors (reviewed in refs 65, 66). Clostridial toxins covalently modify and inactivate Rho GTPases, whereas *Yersinia* YopE (injected into cells through a bacterial type III secretion system) acts as a Rho GAP. In both cases, this prevents phagocytosis of bacteria. *Salmonella typhimurium*, on the other hand, is an invasive bacterium and it injects cells with SopE, which acts as a GEF, to activate Cdc42 and Rac, and thereby promotes its own internalization. Furthermore, once inside the host cell, *Salmonella* releases SptP, a GAP that downregulates Cdc42 and Rac.

Proliferation

In addition to their effects on the cytoskeleton, Rho GTPases contribute to the regulation of cell cycle progression. One of the best-characterized examples of this is after antigen stimulation of T cells. Appropriate engagement of the T-cell receptor induces GTP loading on Rac (catalysed by Vav) and stimulation of the JNK MAP kinase cascade. This leads to activation of the transcription factor NF-AT and expression of the cytokine interleukin-2 (IL-2), which in turn stimulates G1 progression of the quiescent T cell and promotes clonal expansion of the appropriate antigen-specific lymphocytes⁶⁷. Interestingly, Rac has a dual role and controls changes to the actin cytoskeleton that are essential for induction of proliferation in these cells. Rac and its exchange factor Vav are also essential for proliferation of B cells; in this case, Rac regulates expression of the cyclin D2 gene through an unknown pathway⁶⁸.

Work with fibroblasts and epithelial cells in culture has shown

that Rho, Rac and Cdc42 each contribute to G1 cell cycle progression. All three can promote entry into G1 and progression to S phase when expressed in quiescent fibroblasts, whereas inhibition of any of the three blocks serum-induced G1 progression (Fig. 4)⁶⁹. The signal transduction pathways involved are still unclear, but Rho seems to have at least two important roles: to inhibit expression of the cyclin/Cdk inhibitor p21^{Waf1/Cip1}, and to induce cyclin D1 expression in mid-G1 (by promoting sustained activation of extracellular-signal-regulated kinase (ERK) MAP kinase)^{70,71}. In endothelial cells, Rac is essential for cyclin D1 expression during G1, although in this case, it regulates messenger RNA translation rather than transcription of the gene⁷².

The cell cycle is completed with cytokinesis (Fig. 4), and in animal cells this is driven by an actin and myosin contractile ring, which constricts to form the two daughter cells. Inhibition of either Rho or Cdc42 prevents the assembly of the contractile ring in a variety of mammalian cells as well as in *Xenopus* embryos⁷³. Expression of constitutively activated Rho or Cdc42 also blocks cytokinesis, suggesting that cycling between the active and inactive forms is required for function.

Regulated secretion

There are now several examples where Rho GTPases participate in regulated secretory pathways and in some case this seems to be actin-independent. The interaction of the T-cell receptor (TCR) on cytotoxic T cells with a corresponding antigen-presenting cell (APC), for example, induces the re-orientation of the Golgi and MTOC to face the APC interaction site, thereby facilitating the spatially restricted delivery of secretory granules (Fig. 4). MTOC polarization in T cells is Cdc42-dependent—whether Par6 and atypical PKC are involved has not yet been examined⁷⁴. Crosslinking of IgE receptors on circulating mast cells induces histamine and serotonin release from large cytoplasmic granules, and work with permeabilized cells has revealed that Rho, Rac and Cdc42 are required. Although their exact function is not known, it is thought to be independent of the actin cytoskeleton⁷⁵. Finally, insulin stimulation of adipocytes induces translocation of intracellular vesicles containing glucose transporters, and TC10, a relatively uncharacterized member of the Rho family, is required⁷⁶. In this case, a GTPase-dependent re-organization of cortical actin seems to be involved⁷⁷.

Future perspectives

The single most notable feature of Rho GTPases is their participation in so many aspects of cell biology, ranging from the fundamental (for example, cell polarity) to the highly specialized (for example, contraction of vascular smooth muscle cells). The future will no doubt hold many more surprises and examples of cellular control by these proteins.

The analysis of individual signal transduction pathways controlled by Rho GTPases still poses many problems, and with so many GEFs, GAPs and targets it has proved difficult to link specific receptors all the way through to a biological response. Conceptually, it is still unclear how spatially localized activation of GTPases is achieved or, when activated, how GTPases are able to ‘choose’ the correct target(s) from the large number of possibilities present in the same cell. The answer will probably involve positive feedback loops, cooperative signalling pathways, scaffold proteins and intracellular compartmentalization.

Complex changes in cell behaviour are the result of multiple biochemical pathways acting in unison. Combinations of Rho GTPases, each with their numerous targets, are ideally positioned to coordinate this kind of parallel circuitry. Yeast has provided several examples of this, but we are still awaiting clearcut examples in mammalian cells. Changes in the actin cytoskeleton do not occur independently of other cellular responses such as changes in gene transcription, the microtubule cytoskeleton or vesicular transport.

We have seen how a simple Rho GTPase switch can potentially orchestrate this biological complexity, what is needed now is evidence of how and when they actually do. Indeed, it may be that to achieve a better understanding of the molecular basis of cell biology, it will be necessary to shift focus a little, away from trying to identify every component of a favoured signalling pathway, to exploring how pathways cooperate with each other during a defined biological response. □

doi:10.1038/nature01148.

1. Ridley, A. J. & Hall, A. The small GTP-binding protein Rho regulates the assembly of focal adhesions and stress fibers in response to growth factors. *Cell* **70**, 389–399 (1992).
2. Ridley, A. J., Paterson, C. L., Johnston, C. L., Diekmann, D. & Hall, A. The small GTP-binding protein Rac regulates growth factor-induced membrane ruffling. *Cell* **70**, 401–410 (1992).
3. Nobes, C. D. & Hall, A. Rho, Rac and Cdc42 regulate the assembly of multimolecular focal complexes associated with actin stress fibers, lamellipodia and filopodia. *Cell* **81**, 53–62 (1995).
4. Kozma, R., Ahmed, S., Best, A. & Lim, L. The Ras-related protein Cdc42Hs and bradykinin promote formation of peripheral actin microspikes and filopodia in Swiss 3T3 fibroblasts. *Mol. Cell. Biol.* **15**, 1942–1952 (1995).
5. Pruyne, D. & Bretscher, A. Polarization of cell growth in yeast. I. Establishment and maintenance of polarity states. *J. Cell Sci.* **113**, 365–375 (2000).
6. Schmidt, A. & Hall, M. N. Signaling to the actin cytoskeleton. *Annu. Rev. Cell Dev. Biol.* **14**, 305–338 (1998).
7. Butty, A.-C. *et al.* A positive feedback loop stabilizes the guanine-exchange factor Cdc24 at sites of polarization. *EMBO J.* **21**, 1565–1576 (2002).
8. Kemphues, K. PARs and embryonic polarity. *Cell* **101**, 345–348 (2000).
9. Ohno, S. Intercellular junctions and cellular polarity: the PAR-aPKC complex, a conserved core cassette playing fundamental roles in cell polarity. *Curr. Opin. Cell Biol.* **13**, 641–648 (2001).
10. Gotta, M., Abraham, M. C. & Ahringer, J. CDC-42 controls early cell polarity and spindle orientation in *C. elegans*. *Curr. Biol.* **11**, 482–488 (2001).
11. Kay, A. J. & Hunter, C. P. CDC-42 regulates PAR protein localization and function to control cellular and embryonic polarity in *C. elegans*. *Curr. Biol.* **11**, 474–481 (2001).
12. Joberty, G., Petersen, C., Gao, L. & Macara, I. G. The cell-polarity protein Par6 links Par3 and atypical protein kinase C to Cdc42. *Nature Cell Biol.* **2**, 531–539 (2000).
13. Qiu, R.-G., Abo, A. & Martin, G. S. A human homolog of the *C. elegans* polarity determinant Par-6 links Rac and Cdc42 to PKC ζ signaling and cell transformation. *Curr. Biol.* **10**, 697–707 (2000).
14. Lin, D. *et al.* A mammalian PAR-3-PAR-6 complex implicated in Cdc42/Rac1 and aPKC signalling and cell polarity. *Nature Cell Biol.* **2**, 540–547 (2000).
15. Van Aelst, L. & Symons, M. Role of Rho family GTPases in epithelial morphogenesis. *Genes Dev.* **16**, 1032–1054 (2002).
16. Vasioukhin, V., Bauer, C., Yin, M. & Fuchs, E. Directed actin polymerization is the driving force for epithelial cell-cell adhesion. *Cell* **100**, 209–219 (2000).
17. Raich, W. B., Agbunag, C. & Hardin, J. Rapid epithelial-sheet sealing in the *Caenorhabditis elegans* embryo requires cadherin-dependent filopodial priming. *Curr. Biol.* **9**, 1139–1146 (1999).
18. Jacinto, A., Martinez-Arias, A. & Martin, P. Mechanisms of epithelial fusion and repair. *Nature Cell Biol.* **3**, E117–E123 (2001).
19. Yamanaka, T. *et al.* PAR-6 regulates aPKC activity in a novel way and mediates cell-cell contact-independent formation of the epithelial junctional complex. *Genes Cells* **6**, 721–731 (2001).
20. Rojas, R., Ruiz, W. G., Leung, S. M., Jou, T. S. & Apodaca, G. Cdc42-dependent modulation of tight junctions and membrane protein traffic in polarized Madin-Darby canine kidney cells. *Mol. Biol. Cell* **12**, 2257–2274 (2001).
21. Gao, L., Joberty, G. & Macara, I. G. Assembly of epithelial tight junctions is negatively regulated by Par-6. *Curr. Biol.* **12**, 221–225 (2002).
22. Ebneth, K. *et al.* The cell polarity protein ASIP/Par-3 directly associates with junctional adhesion molecule (JAM). *EMBO J.* **20**, 3738–3748 (2001).
23. O'Brien, L. E. *et al.* Rac1 orientates epithelial apical polarity through effects on basolateral laminin assembly. *Nature Cell Biol.* **3**, 831–838 (2001).
24. Kroschewski, R., Hall, A. & Mellman, I. Cdc42 controls secretory and endocytic transport to the basolateral plasma membrane of MDCK cells. *Nature Cell Biol.* **1**, 8–13 (1999).
25. Zhang, X. *et al.* Cdc42 interacts with the exocyst and regulates polarized secretion. *J. Biol. Chem.* **276**, 46745–46750 (2001).
26. Sugihara, K. *et al.* The exocyst complex binds the small GTPase RalA to mediate filopodia formation. *Nature Cell Biol.* **4**, 73–78 (2001).
27. Moskalenko, S. *et al.* The exocyst is a Ral effector complex. *Nature Cell Biol.* **4**, 66–72 (2002).
28. Luo, L. Rho GTPases in neuronal morphogenesis. *Nature Rev. Neurosci.* **1**, 173–180 (2000).
29. Lee, T., Winter, C., Marticic, S. S., Lee, A. & Luo, L. Essential role of *Drosophila* RhoA in the regulation of neuroblast proliferation and dendritic but not axonal morphogenesis. *Neuron* **25**, 307–316 (2000).
30. Ozdinler, P. H. & Erzurumlu, R. S. Regulation of neurotrophin-induced axonal responses via Rho GTPases. *J. Comp. Neurol.* **438**, 377–387 (2001).
31. Li, Z., Van Aelst, L. & Cline, H. T. Rho GTPases regulate distinct aspects of dendritic arbor growth in *Xenopus* central neurons *in vivo*. *Nature Neurosci.* **3**, 217–225 (2000).
32. Wong, W. T., Faulkner-Jones, B., Sanes, J. R. & Wong, R. O. Rapid dendritic remodeling in the developing retina: dependence on neurotransmission and reciprocal regulation by Rac and Rho. *J. Neurosci.* **20**, 5024–5036 (2000).
33. Ridley, A. Rho GTPases and cell migration. *J. Cell Sci.* **114**, 2713–2722 (2001).
34. Small, J. V., Stradal, T., Vignal, E. & Rottler, K. The lamellipodium: where motility begins. *Trends Cell Biol.* **12**, 112–120 (2002).
35. Condeelis, J. How is actin polymerization nucleated *in vivo*? *Trends Cell Biol.* **11**, 288–293 (2001).
36. Kraynov, V. S. *et al.* Localized Rac activation dynamics visualized in living cells. *Science* **290**, 333–337 (2000).

37. Gupta, T. & Schupbach, T. Two signals are better than one: border cell migration in *Drosophila*. *Dev. Cell* **1**, 443–445 (2001).
38. Ducheck, P., Somogyi, K., Jekely, G., Beccari, S. & Rorth, P. Guidance of cell migration by the *Drosophila* PDGF/VEGF receptor. *Cell* **107**, 17–26 (2001).
39. Reddien, P. W. & Horvitz, H. R. CED-2/CrkII and CED-10/Rac control phagocytosis and cell migration in *Caenorhabditis elegans*. *Nature Cell Biol.* **2**, 131–136 (2000).
40. Albert, M. L., Kim, J. I. & Birge, R. B. α v β 5 integrin recruits the CrkII-Dock180-rac1 complex for phagocytosis of apoptotic cells. *Nature Cell Biol.* **2**, 899–905 (2000).
41. Worthylake, R. A., Lemoine, S., Watson, J. M. & Burridge, K. RhoA is required for monocyte tail retraction during transendothelial migration. *J. Cell Biol.* **154**, 147–160 (2001).
42. Allen, W. E., Zicha, D., Ridley, A. J. & Jones, G. E. A role for Cdc42 in macrophage chemotaxis. *J. Cell Biol.* **141**, 1147–1157 (1998).
43. Wittmann, T. & Waterman-Storer, C. M. Cell motility: can Rho GTPases and microtubules point the way? *J. Cell Sci.* **114**, 3795–3803 (2001).
44. Liu, B. P. & Strittmatter, S. M. Semaphorin-mediated axonal guidance via Rho-related G proteins. *Curr. Opin. Cell Biol.* **13**, 619–626 (2001).
45. Ng, J. *et al.* Rac GTPases control axon growth, guidance and branching. *Nature* **416**, 442–447 (2002).
46. Lundquist, E. A., Reddien, P. W., Hartwig, E., Horvitz, H. R. & Bargmann, C. I. Three *C. elegans* Rac proteins and several alternative Rac regulators control axon guidance, cell migration and apoptotic cell phagocytosis. *Development* **128**, 4475–4488 (2001).
47. Brock, J., Midwinter, K., Lewis, J. & Martin, P. Healing of incisional wounds in the embryonic chick wing bud: characterization of the actin purse-string and demonstration of a requirement for Rho activation. *J. Cell Biol.* **135**, 1097–1107 (1996).
48. Lu, Y. & Settleman, J. The *Drosophila* Pkn protein kinase is a Rho/Rac effector target required for dorsal closure during embryogenesis. *Genes Dev.* **13**, 1168–1180 (1999).
49. Stronach, B. E. & Perrimon, N. Stress signaling in *Drosophila*. *Oncogene* **18**, 6172–6182 (1999).
50. Knust, E. *Drosophila* morphogenesis: movements behind the edge. *Curr. Biol.* **7**, R558–R561 (1997).
51. Nobes, C. D. & Hall, A. Rho GTPases control polarity, protrusion and adhesion during cell movement. *J. Cell Biol.* **144**, 1235–1244 (1999).
52. Etienne-Manneville, S. & Hall, A. Integrin-mediated Cdc42 activation controls cell polarity in migrating astrocytes through PKC ζ . *Cell* **106**, 489–498 (2001).
53. Palazzo, A. F. *et al.* Cdc42, dynein, and dynactin regulate MTOC reorientation independent of Rho-regulated microtubule stabilization. *Curr. Biol.* **11**, 1536–1541 (2001).
54. Ishizaki, T. *et al.* Coordination of microtubules and actin cytoskeleton by the Rho effector mDia. *Nature Cell Biol.* **3**, 8–14 (2001).
55. Ridely, A. J. Rho family proteins: coordinating cell responses. *Trends Cell Biol.* **11**, 471–477 (2001).
56. Fukata, Y., Amano, M. & Kaibuchi, K. Rho-Rho-kinase pathway in smooth muscle contraction and cytoskeletal reorganization of non-muscle cells. *Trends Pharmacol. Sci.* **22**, 32–39 (2001).
57. Sakurada, S., Okamoto, H., Takuwa, N., Sugimoto, N. & Takuwa, Y. Rho activation in excitatory agonist-stimulated vascular smooth muscle. *Am. J. Physiol. Cell Physiol.* **281**, C571–C578 (2001).
58. Uehata, M. *et al.* Calcium sensitization of smooth muscle mediated by Rho-associated protein kinase in hypertension. *Nature* **389**, 990–994 (1997).
59. van Nieuw Amerongen, G. P., van Delft, S., Vermeer, M. A., Collard, J. G. & van Hinsbergh, V. W. Activation of RhoA by thrombin in endothelial hypermeability: role of Rho kinase and protein tyrosine kinases. *Circ. Res.* **87**, 335–340 (2000).
60. Wojciak-Stothard, B., Potempa, S., Eichholtz, T. & Ridley, A. J. Rho and Rac but not Cdc42 regulate endothelial cell permeability. *J. Cell Sci.* **114**, 1343–1355 (2001).
61. Caron, E. & Hall, A. Identification of two distinct mechanisms of phagocytosis controlled by different Rho GTPases. *Science* **282**, 1717–1721 (1998).
62. Caron, E. Rac and roll over the corpses. *Curr. Biol.* **10**, R489–R491 (2000).
63. Bokoch, G. M. Regulation of cell function by Rho family GTPases. *Immunol. Res.* **21**, 139–148 (2000).
64. Nobes, C. & Marsh, M. Dendritic cells: new roles for Cdc42 and Rac in antigen uptake? *Curr. Biol.* **10**, R739–R741 (2000).
65. Aktories, K., Schmidt, G. & Just, I. Rho GTPases as targets of bacterial protein toxins. *Biol. Chem.* **381**, 421–426 (2000).
66. Galan, J. E. *Salmonella* interactions with host cells: Type III secretion at work. *Annu. Rev. Cell Dev. Biol.* **17**, 53–86 (2001).
67. Cantrell, D. Lymphocyte signalling: a coordinating role for Vav? *Curr. Biol.* **8**, R535–R538 (1998).
68. Glassford, J. *et al.* Vav is required for cyclin D2 induction and proliferation of mouse B lymphocytes activated via the antigen receptor. *J. Biol. Chem.* **276**, 41040–41048 (2001).
69. Olson, M. F., Ashworth, A. & Hall, A. An essential role for Rho, Rac and CDC42 GTPases in cell cycle progression through G1. *Science* **269**, 1270–1272 (1995).
70. Welsh, C. F. *et al.* Timing of cyclin D1 expression within G1 phase is controlled by Rho. *Nature Cell Biol.* **3**, 950–957 (2001).
71. Olson, M. F., Paterson, H. F. & Marshall, C. J. Signals from Ras and Rho GTPases interact to regulate expression of p21^{Waf1/Cip1}. *Nature* **394**, 295–299 (1998).
72. Mettouchi, A. *et al.* Integrin-specific activation of Rac controls progression through the G1 phase of the cell cycle. *Mol. Cell* **8**, 115–127 (2001).
73. Glotzer, M. Animal cell cytokinesis. *Annu. Rev. Cell Dev. Biol.* **17**, 351–386 (2001).
74. Stowers, L., Yelon, D., Berg, L. J. & Chant, J. Regulation of the polarization of T cells towards antigen-presenting cells by Ras-related GTPase CDC42. *Proc. Natl Acad. Sci. USA* **92**, 5027–5031 (1995).
75. Pinxteren, J. A., O'Sullivan, A. J., Larbi, K. Y., Tatham, P. E. R. & Gomperts, B. D. Thirty years of stimulus-secretion coupling: from Ca²⁺ to GTP in the regulation of exocytosis. *Biochimie* **82**, 385–393 (2000).
76. Chiang, S. H. *et al.* Insulin-stimulated GLUT4 translocation requires the CAP-dependent activation of TC10. *Nature* **410**, 944–948 (2001).
77. Kanzaki, M. & Pessin, J. E. Insulin-stimulated GLUT4 translocation in adipocytes is dependent upon cortical actin remodeling. *J. Biol. Chem.* **276**, 42436–42444 (2001).

Acknowledgements We are grateful for support from Cancer Research UK and the Medical Research Council. S.E.-M. is supported by an EMBO long-term fellowship.

Correspondence and requests for materials should be addressed to A.H. (e-mail: alan.hall@ucl.ac.uk).

Structural basis of BMP signalling inhibition by the cystine knot protein Noggin

Jay Groppe*†, Jason Greenwald*, Ezra Wiater‡, Joaquin Rodriguez-Leon§, Aris N. Economides||, Witek Kwiatkowski*, Markus Affolter†, Wylie W. Vale‡, Juan Carlos Izpisua Belmonte§¶ & Senyon Choe*

* Structural Biology, ‡ Peptide Biology, and ¶ Gene Expression Laboratories, Salk Institute, 10010 North Torrey Pines Road, La Jolla, California 92037, USA

§ Instituto Gulbenkian de Ciencia, Rua da Quinta Grande, Oeiras 2780-901, Portugal

† Biozentrum, Klingelbergstrasse 50-70, Basel CH-4056, Switzerland

|| Regeneron Pharmaceuticals, 777 Old Saw Mill River Road, Tarrytown, New York, New York 10591, USA

The interplay between bone morphogenetic proteins (BMPs) and their antagonists governs developmental and cellular processes as diverse as establishment of the embryonic dorsal–ventral axis, induction of neural tissue, formation of joints in the skeletal system and neurogenesis in the adult brain. So far, the three-dimensional structures of BMP antagonists and the structural basis for inactivation have remained unknown. Here we report the crystal structure of the antagonist Noggin bound to BMP-7, which shows that Noggin inhibits BMP signalling by blocking the molecular interfaces of the binding epitopes for both type I and type II receptors. The BMP-7-binding affinity of site-specific variants of Noggin is correlated with alterations in bone formation and apoptosis in chick limb development, showing that Noggin functions by sequestering its ligand in an inactive complex. The scaffold of Noggin contains a cystine (the oxidized form of cysteine) knot topology similar to that of BMPs; thus, ligand and antagonist seem to have evolved from a common ancestral gene.

The generation of overall body plans and the formation of organs in animal embryos depend on intercellular signalling pathways throughout development. This cell–cell communication is mediated primarily by secreted protein ligands such as Notch, Hedgehog, Wntless, fibroblast growth factor (FGF), epidermal growth factor (EGF) and transforming growth factor- β (TGF- β) and their respective cell-surface receptors. The TGF- β superfamily includes TGF- β , Activin, Nodal, muellerian inhibiting substance (MIS), growth and differentiation factor (GDF) and BMP families. Signal transduction by these ligands is modulated through overlapping receptor specificity¹, by temporal and tissue-specific regulation of gene expression² and by secreted, high-affinity binding proteins that sequester the ligands and interrupt the signal³.

BMPs, which were originally identified from protein extracts of demineralized bone⁴, are involved in the complete cascade of body patterning and morphogenesis⁵. This pathway is one of the evolutionarily oldest used by animal embryos and is required for dorsal–ventral patterning of the early embryo of both vertebrates and invertebrates². In addition to tissue-specific expression of the ligands and their cell-surface receptors, BMP signalling is regulated by negative-acting factors to pattern correctly fields of cells and shape structures. This antagonism is accomplished primarily by three classes of secreted inhibitory protein: Noggin⁵; the DAN family⁶, which includes DAN, Cerberus and Gremlin; and vertebrate Chordin⁷ and *Drosophila* SOG.

Noggin is a BMP-specific antagonist protein found to rescue dorsal (head) structures in ventralized *Xenopus* embryos⁷. In regulation of vertebrate dorsal–ventral patterning and later skeletogenesis, the primary physiological role of Noggin is to antagonize the action of BMPs directly^{8,9}. In contrast to the normal appearance of mice heterozygous for a null mutation in the Noggin gene⁹, heterozygous mutations in the human locus result in individuals with apical joint fusions, indicative of functional haploinsufficiency^{10–12}. Antagonism by Noggin has been implicated as the primary inducer of neuronal differentiation in stem cells of the adult subventricular zone of the brain. Blocking the inhibitory BMP

signal provides a neurogenic microenvironment that allows the proliferation of neuroblasts¹³.

Human Noggin protein

Human Noggin (hNoggin) has 205 amino acids (residues 28–232) after removal of its signal sequence (residues 1–27) and is secreted as a glycosylated, covalently linked homodimer (Fig. 1). Although a Noggin-like gene has been described in the flatworm *Dugesia*¹⁴, homologues have been identified principally in vertebrates with a single copy per genome except for zebrafish (three) and pufferfish (two). The primary structure of Noggin consists of an acidic amino-terminal half and a cysteine-rich (nine cysteine residues) carboxy-terminal half, for which no apparent structural homology has been detected. A central, highly basic heparin-binding segment retains Noggin at the cell surface, thereby regulating its diffusion^{15,16}. hNoggin binds with high affinity to hBMP-4 (and with lower affinity to hBMP-7)⁸ and binds comparably to homologues from *Drosophila* and *Xenopus*¹⁷, emphasizing its high degree of functional conservation.

Structure determination

The protein complex (relative molecular mass 80,000; M_r 80K) of one dimer of hNoggin and one dimer of hBMP-7 was isolated by size-exclusion chromatography and crystallized at neutral pH. The structure was solved at 2.4 Å resolution by the multiwavelength anomalous diffraction (MAD) method after rapidly soaking a crystal in a cryosolvent containing NaBr¹⁸. The anomalous signal from ten bound bromide ions provided sufficient phasing for a readily interpretable electron density map (Supplementary Table 1). Model building was facilitated by placement of a free BMP-7 dimer¹⁹ in the map and by independent identification of cystine and methionine sulphur positions from their anomalous signals.

Structure of the Noggin–BMP-7 complex

The elongated Noggin homodimer shares a common two-fold symmetry axis with BMP-7, creating a large ringed structure

(Fig. 2a). BMP-7 is a butterfly-shaped homodimer, each wing of which is formed by two pairs of antiparallel β -strands (referred to as fingers 1 and 2) and extends from a core containing seven disulphide bonds. The curvature of the extended fingers creates a concave side to which $\alpha 3$ of the other subunit binds, stabilizing the dimer. The core of BMP-7 is characterized by its trademark eight-membered ring, comprising segments of five (CEGEC₇₁) and three (CGC₁₃₈) residues linked by bonds between Cys 67 (II) and Cys 136 (V), and between Cys 71 (III) and Cys 138 (VI). A third disulphide bond between Cys 38 (I) and Cys 104 (IV) penetrates through the middle of the ring, forming a ten-membered cystine knot²⁰.

The topology of the C-terminal half of Noggin resembles BMPs, in that two pairs of β -strands ($\beta 1$ – $\beta 2$, 'finger 1'; $\beta 3$ – $\beta 4$, 'finger 2') extend out from a core containing several disulphide bonds. This cystine-rich scaffold is preceded by an N-terminal segment of about 20 residues, which we refer to as the 'clip' (residues 28–48), and a 'helical' domain of about 100 residues that extends from near the tip of finger 2 to the dimer interface (residues 49–154). Noggin is most distinct from BMP-7 in the way the two subunits bind one another; that is, a helix equivalent to $\alpha 3$ of BMP-7 (Figs 1 and 2a) is absent, and the long $\alpha 4$ of the helical domain provides an additional dimer interaction. Through the head-to-head (the 'head' being the disulphide-bonded core) dimerization interface, the whole molecule can be described as a 'clamp-shaped' dimer with two extended 'clip' segments.

Noggin binding induces a significant conformational change in BMP-7 (Fig. 2b). The tips of fingers 1 and 2 of BMP-7 curl around the N-terminal segment of Noggin, pivoting off the tip (3_{10} helix) of finger 1 with the largest movement in finger 2. This marked change in conformation contrasts with less-pronounced changes in BMP-2 and BMP-7 bound to receptor extracellular domains^{21,22}. Portions of the helical domain of Noggin are relatively mobile, particularly in the region near the tip of finger 2 and in the heparin-binding loop

between $\alpha 3$ and $\alpha 4$ (Fig. 2c). Removal of this loop (Lys133–Lys144) does not affect the affinity of Noggin for BMP¹⁵. The most flexible part of Noggin is the polyglycine loop (GGGGGAA₉₅), which is crystallographically invisible and present only in mammalian Noggins.

Noggin is related to cystine knot growth factors

As in BMP-7, three disulphides of the Noggin monomer are in a cystine knot topology (Supplementary Fig. 1). But the knot is atypical with respect to that of the ligand and those proposed for the DAN family of antagonists²³ in that the CXGXC motif of ten-membered knots²⁰ is extended to seven residues (CX₅C) in Noggin and lacks the central conserved glycine, resulting in a noncanonical, 12-membered cystine knot (10-membered ring). The connectivities of the knots are identical, however, with two disulphide bonds between positions II and V, and positions III and VI, and the through-ring bond between positions I and IV (Fig. 1).

Although the scaffolds of the cystine knot domains are highly similar (Fig. 3), the head-to-head dimerization mode of Noggin differs markedly from the back-to-back (nerve growth factor)²⁴ and head-to-tail (TGF- β , platelet-derived growth factor, human chorionic gonadotropin) modes observed in the cystine knot growth factor superfamily²⁵. Despite this difference, the buried surface areas of the Noggin (1,399 Å²) and BMP-7 (1,328 Å²) dimers are comparable.

Basis for inhibition of receptor binding by Noggin

The surfaces of BMP ligands have two prominent hydrophobic patches¹⁷, one on the convex type II receptor-binding interface (ref. 21 and Fig. 2c, box II) and the other on the concave type I receptor-binding interface (ref. 22 and Fig. 2c, box I). Superposition of the Noggin–BMP-7 structure onto a model of the BMP signalling complex^{21,22} shows that Noggin binding effectively masks both pairs

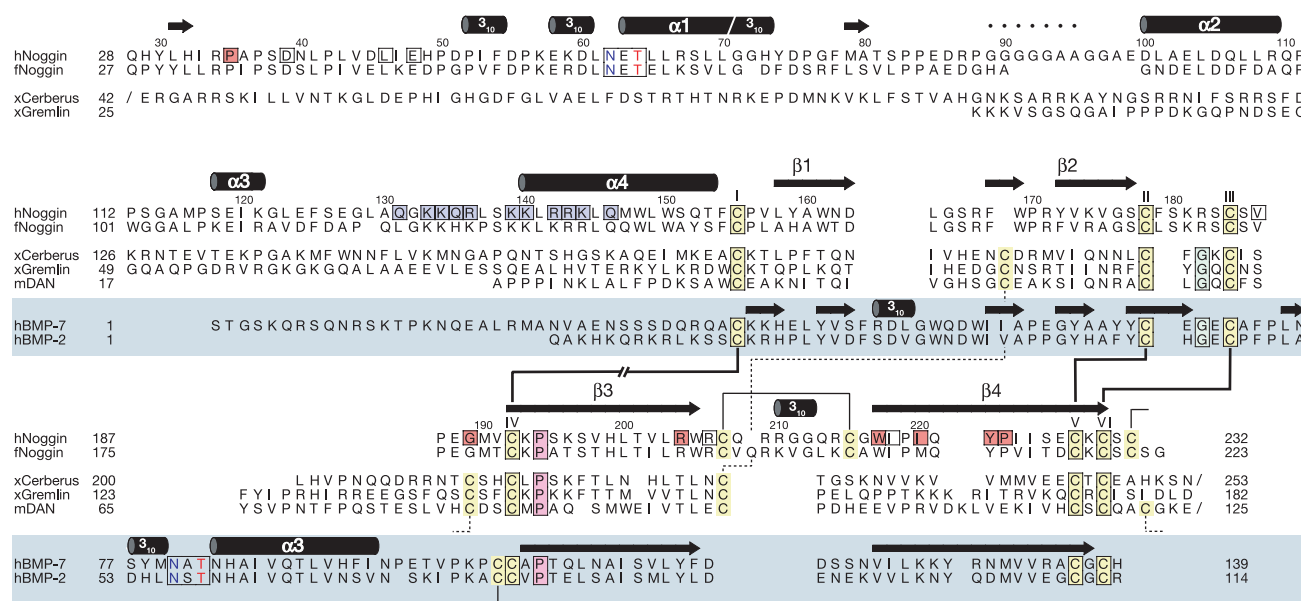


Figure 1 Sequence alignment of Noggin, DAN family antagonists and BMPs. Human and Fugu (*Takifugu rubripes*) Noggins, mouse DAN, *Xenopus* Gremlin and Cerberus and human BMP-7 and BMP-2 protein sequences were aligned by CLUSTAL W⁴⁸ with a cystine knot profile by STAMP⁴⁹. Conserved cysteines of cystine knots, labelled I to VI, are linked by thick lines (the 'through-ring' bond is indicated by '— // —'). The disulphide in finger 2 of Noggin (Cys207–Cys215) is indicated by a thin unbroken line, intermolecular disulphide bonds are indicated by short right angles (half-cystines), and predicted linkages of DAN proteins are indicated by thin dashed lines. Other features are indicated

as follows: mutations in hNoggin are boxed in red, variants are in open boxes, asparagines of glycosylation sites of BMPs and Noggin are blue, threonines of boxed NXT signals are red, the heparin-binding site of hNoggin is boxed in blue, glycine in CXGXC motif of ten-membered cystine knot proteins are boxed in green, the conserved proline flanking the knots is boxed in pink. N- and C-terminal sequences of secreted Cerberus, truncated by slash marks, contain an additional 22 and 17 residues, respectively, and the C-terminal sequence of DAN contains an additional 53 residues.

of binding epitopes (Fig. 4a). This is consistent with the known ability of Noggin to inhibit binding to both types of receptor⁸.

The type I receptor-binding site is occluded by a segment of the clip domain (Gln 28 to Asp 39), which is bound predominantly through hydrophobic interactions (Fig. 4b, box I). The hydrophobic ring of Pro 35 of Noggin inserts into a hydrophobic pocket on BMP-7 formed by Trp 52, Trp 55, Val 87, Tyr 128 and Met 131, mimicking a similar insertion of Phe 85 from $\alpha 1$ of the BMP receptor BMPRIa into the hydrophobic cleft on BMP-2 (ref. 22). This residue is thought to be a key determinant of all type I receptors. Notably, the short $\alpha 1$ helix seems to be unstructured in free BMPRIa²⁶. Because residues lining this cleft are highly conserved between BMPs, but less so among other TGF- β ligands, interactions at this site may also provide a basis for Noggin's high affinity and specificity for BMP. Additional binding specificity may arise from the pre-helix loop of BMP, which varies in length among related TGF- β subfamilies.

The type II receptor-binding site is masked extensively by the C-terminal half of the clip segment (Asn 40 to Glu 48), by the distal tip of finger 1, and by finger 2 (Fig. 4c, box II). As in box I, the interactions highlighted in box II are also predominantly hydro-

phobic. The extra disulphide bond present near the tip of finger 2 (Cys207–Cys215), which is unique to Noggin, enhances the conformational rigidity of the long finger.

Thus, the two means by which Noggin binds BMP-7 have marked resemblances to those of the receptors: first, a hydrophobic side chain from a flexible backbone segment is inserted into the hydrophobic pocket of BMP; and second, complementary interactions occur between two curved hydrophobic surfaces. An additional resemblance is that Noggin, as a homodimer, binds to both pairs of receptor-binding epitopes of BMP-7, mimicking the binding mode of the heterotetrameric receptor assembly. Although the buried surface area for the Noggin–BMP-7 interface (1,444 Å² per subunit) is less than the combined surface areas buried by the type I and type II receptors ($\sim 1,800$ Å²)^{21,22}, the overall affinity of Noggin should be enhanced by the proximity effect owing to its dimeric nature.

Binding affinity of Noggin variants *in vitro*

To probe the molecular interface between BMP-7 and Noggin, we chose eight positions that show varying degrees of alteration in binding affinity without compromising the folding efficiency or

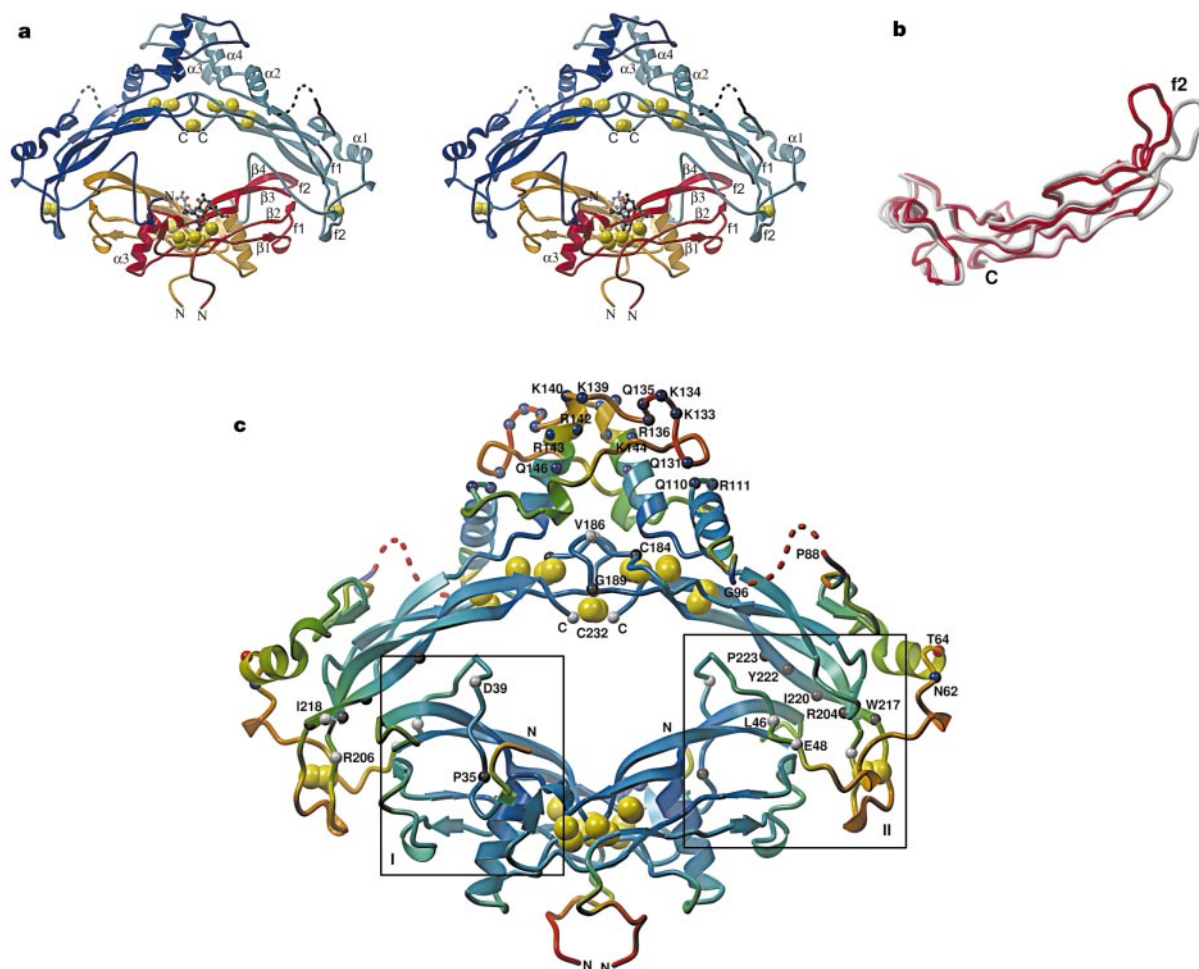


Figure 2 Structure of the Noggin–BMP-7 complex. **a**, Ribbon diagram of the Noggin–BMP-7 complex in stereo, viewed with the two-fold symmetry axis vertical. Noggin subunits are blue and light blue, BMP-7 homodimer subunits are red and gold, sulphur atoms of cystine linkages are yellow spheres, and the N-acetylglucosamines of BMP-7 are shown as balls-and-sticks. The fingers (f1 and f2) of Noggin and BMP-7 are labelled near their tips. **b**, Superposition of free (grey) and Noggin-bound (red) BMP-7 monomers, viewed from above the complex shown in **a** down the two-fold axis. **c**, Noggin–BMP-7

complex coloured as a rainbow gradient with respect to B-factor values (blue, lowest; red, highest) to depict relative order and disorder. Substitutions in disease-causing Noggin mutants are black, substitutions in variant constructs are light grey (Asp 39, Leu 46, Glu 48, Val 186, Arg 206, Ile 218, Cys 232) or black (Pro 35). Heparin-binding site residues of Noggin (Gln 131 to Lys 144 and putatively Gln 110, Arg 111) are blue. Asn 62 and Thr 64 comprise the N-linked glycosylation signal. Only one set of residues is labelled, although both chains are marked. Boxed regions I and II frame the views in Fig. 4b, c, respectively.

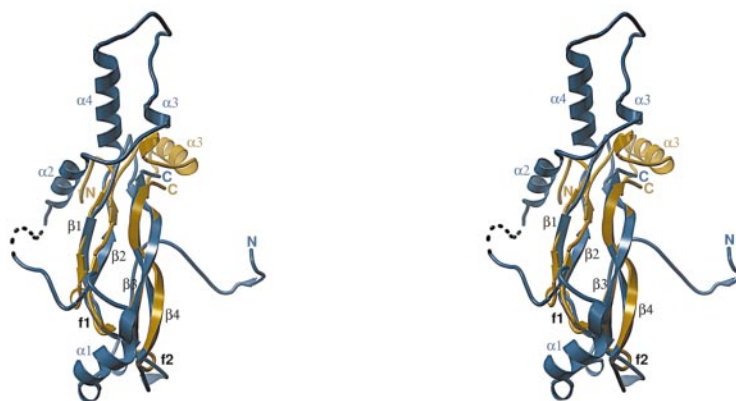


Figure 3 The Noggin monomer contains a 'growth factor' cystine knot scaffold. Superposition of Noggin (blue) and free BMP-7 (gold) monomers. Common β -strands ($\beta 1$ – $\beta 4$) for fingers f1 and f2 are labelled. In the conserved cystine knot scaffold, 67 C α

atoms align with an r.m.s. deviation of 1.74 Å, which is in the range (1.4–3.7 Å) observed for the cystine knot growth factor superfamily²⁵.

stability, and generated point mutations (Pro35Arg, Asp39Ala, Leu46Asp, Glu48Lys, Val186Ala, Arg206Ala, Ile218Glu, Cys232Ser). Three hydrophobic side chains, Pro 35, Leu 46 and Ile 218, that make hydrophobic contacts with BMP-7 were replaced with charged residues that would preclude close contact between the two surfaces. Asp 39 and Glu 48 form ion pairs with Lys 126 of BMP-7 (Fig. 4b, box I) or with Arg 204 of Noggin (Fig. 4c, box II), respectively. Arg 206 lies on the periphery of the box II interface. Two other residues are located in noncovalent (Val 186) and covalent (Cys 232) parts of the Noggin dimer interface.

Relative to wild-type Noggin (effector concentration for half-maximum response (EC₅₀) of 120 pM), the BMP-7 binding affinities of the variants could be grouped into three categories (Fig. 5): first, unperturbed, wild-type-like (Asp39Ala, 220 pM; Val186Ala, 180 pM; Arg206Ala, 95 pM; Cys232Ser, 250 pM); second, unquantifiably low, essentially abolished (Leu46Asp, Glu48Lys, Ile218Glu); and third, diminished (Pro35Arg, 830 pM). Notably, elimination of the intermolecular ion pair (Asp39Ala) did not significantly alter the affinity, whereas disruption of the adjacent hydrophobic interaction (Pro35Arg) reduced binding by about sevenfold. Binding was unaltered with Arg206Ala, but abolished with the charge-substituted Leu46Asp and Ile218Glu variants and the charge-inverted Glu48Lys variant. Neither disruption of the covalent linkage (Cys232Ser) nor destabilization (Val186Ala) of the dimer interface affected binding, indicating that the noncovalent interactions between monomers are both extensive and sufficient for stabilizing the dimer.

In vivo activity in the chick limb bud

The relationship between binding affinities of Noggin and its biological activity was addressed with the chick limb bud, focusing on the outgrowth of the autopod. BMPs, along with their antagonists Noggin^{9,27,28}, Gremlin^{29–31}, Follistatin³² and Chordin³³, are key factors controlling chondrogenesis of digits^{28,34–36} and interdigital apoptosis^{37–39}. For analysis of these processes, beads soaked with purified Noggin were applied directly to the developing chick limb bud (Fig. 6). We found that the progression of limb bud growth is directly correlated with the range of binding activities of Noggin (Fig. 6a–o). Wild-type Noggin and the Cys232Ser variant completely inhibited chondrogenesis (digits truncated at first phalanx), whereas the reduced-binding variant Pro35Arg inhibited chondrogenesis only weakly. After a longer period, Pro35Arg eventually inhibited joint formation (Fig. 6e). The binding-defective Glu48Lys, Ile218Glu and Leu46Asp variants failed to produce any alteration in chondrogenic aggregation. Similarly, expression of the *sox9* gene⁴⁰, an early marker of chondrogenesis, was inhibited by wild-type and Cys232Ser, compromised by Pro35Arg, and unaffected by Glu48Lys, Ile218Glu and Leu46Asp Noggin.

Interdigital apoptosis, another key process of limb development induced by BMPs, was also inhibited in a consistent manner (Fig. 6p–v): that is, completely (wild type, Cys232Ser), slightly (Pro35Arg), or not at all (Asp39Ala, Ile218Glu, Leu46Asp). In addition, sections of treated interdigital spaces were processed by TdT-mediated dUTP nick end labelling (TUNEL) assay to detect DNA degradation as a marker for apoptosis (Fig. 6w–z). Again, both wild-type and Cys232Ser Noggin inhibited apoptosis, whereas Pro35Arg did so weakly.

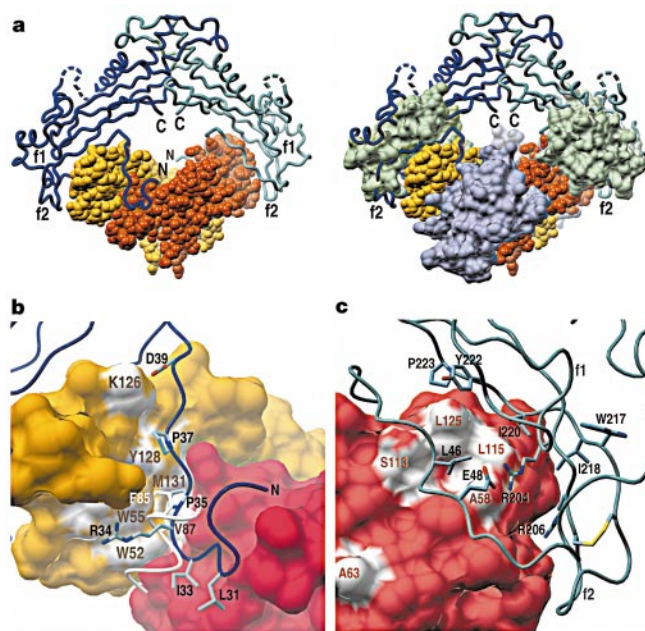


Figure 4 Structural basis for inhibition of receptor binding by Noggin. **a**, Noggin–BMP-7 complex (left) is superimposed on a model of the BMP-7–ActRII–BMPRIa complex (right). BMP-7 monomers (space-filling) are coloured as in Fig. 2a, the pairs of extracellular domains of ActRII and BMPRIa are green and blue, respectively. Note that the fingers (f1 and f2) and the C-terminal half of the 'clip' segment of Noggin interfere with ActRII binding. The N-terminal half of the segment precludes binding by BMPRIa.

b, Surface of BMP-7 with Noggin (blue coil) in the BMP type I receptor-binding epitope (Fig. 2c, box I). Also shown are $\alpha 3$ (white coil) and Phe 85 of BMPRIa from the BMPRIa–BMP-2 complex. **c**, Interactions in the type II receptor-binding epitope (Fig. 2c, box II). Ala 58, Ala 63, Ser 113, Leu 115 and Leu 125 of the BMP type II receptor-binding epitope²¹ are white. Noggin is in light blue, and residues substituted in human mutants and in variant constructs are labelled.

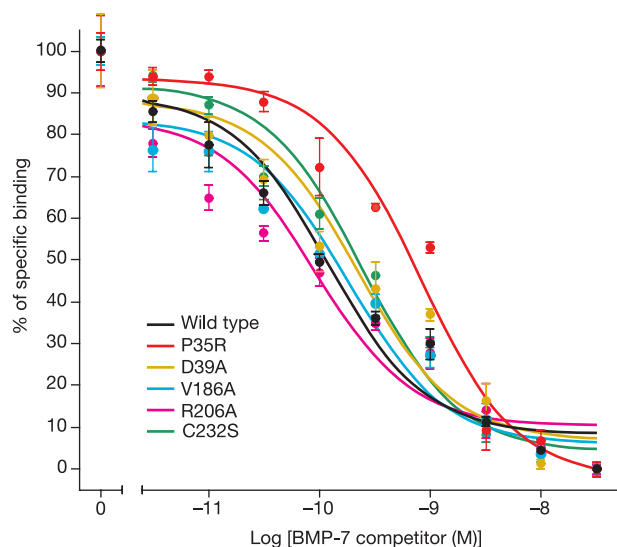


Figure 5 Binding affinity of Noggin proteins for BMP-7. PEG-biotinylated Noggin proteins were incubated with ^{125}I -labelled BMP-7 either alone (100% specific binding) or with increasing concentrations of cold BMP-7 competitor. ^{125}I -labelled BMP-7 binding was normalized, and EC_{50} values were derived for all Noggin proteins except three (Leu46Asp, Glu48Lys and Ile218Glu mutants), which bound negligibly and could not be quantified.

Thus, extending the results of previous assays with blocking antibodies⁸, our analyses show that the activity of Noggin *in vivo* is well-correlated with its ability to bind BMP. In addition, the complete absence of any discernable developmental defects with the three binding-defective Noggin variants, which were otherwise intact, supports the notion that Noggin does not function as a

signal ligand in its own pathway, but instead acts, at least in the experimental setting used, only by sequestration of BMP in an inactive state.

Disease-causing Noggin mutations

Heterozygous missense mutations in the human Noggin gene (*NOG*) are linked to three similar but clinically distinct skeletal dysplasias that result in apical joint fusions^{10–12}. Substitutions of six positions (Pro35Arg, Cys184Tyr, Gly189Cys, Ile220Asn, Tyr222Cys/Tyr222Asn, Pro223Leu) are associated with proximal symphalangism (SYM1), three (Arg204Leu, Pro35Arg, Tyr222Cys) with tarsal/carpal coalition syndrome (TCC), and one (Trp217Gly) with multiple synostoses syndrome (SYNS1). With the exceptions of Arg204Leu (which would disrupt the ion pair with Glu 48) and Pro35Arg, all of the substitutions seem to affect folding and stability rather than BMP-binding affinity (Figs 2c and 4c). Indeed, three mutants have been expressed in transiently transfected COS-7 cells and found to be secreted in reduced (SYM1 mutants, Gly189Cys and Pro223Leu) or undetectable (more severe SYNS1 mutant, Trp217Gly) quantities⁴¹.

Noggin Pro35Arg is particularly interesting because its decreased affinity of only sevenfold leads to diminished inhibition of chondrogenesis. A Pro35Ser mutation that has been identified in an Italian family with symphalangism⁴² probably leads to a comparable decrease in binding affinity, because the polar hydroxyl group of serine would, like the charged guanido group of arginine, be unable to bind the hydrophobic pocket on BMPs in a “knob-into-hole”²² manner (Fig. 4b).

Evolution of regulators of BMP signalling

Noggin serves as a structural model for the other members of the cystine knot family of BMP antagonists. The lack of a single large

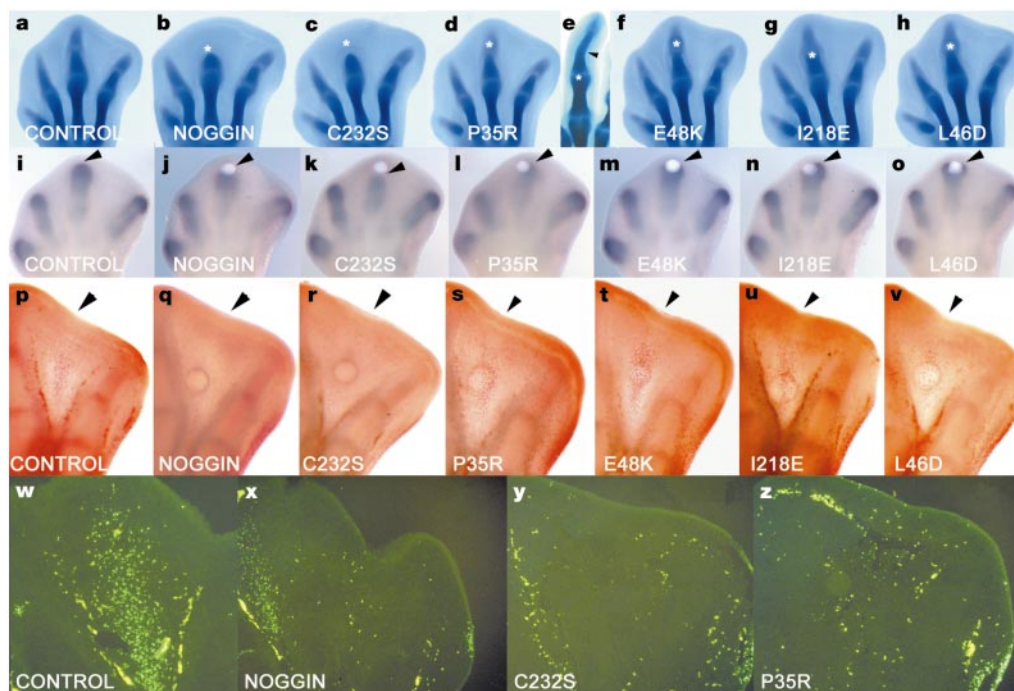


Figure 6 *In vivo* effects of Noggin on BMP-induced chondrogenesis and apoptosis. **a–o**, Implantation of beads soaked in solutions of Noggin was made at the tip of digit III (asterisks). **a–h**, Alcian green staining. Arrowhead in **e** shows digit fusion. **i–o**, RNA *in situ* hybridization with *sox9*. Compare the distance between the tip of labelled digit and the distal part of limb bud (arrowheads) in control and treated limbs. **p–z**, Implantation in the third interdigital space. **p–v**, Vital staining with neutral red. Note that in limbs treated

with wild-type and Cys232Ser Noggin, no indentation is observed at the distal tip of the interdigital space compared with control limbs or limbs treated with Glu48Lys and Ile218Glu and Leu46Asp Noggin (arrowheads). **w–z**, TUNEL labelling in interdigital spaces. The reduced number of fluorescent cells is indicative of reduced cell death in the interdigits of treated versus control limbs.

helix equivalent to $\alpha 3$ of BMP and the positions of cysteines of the DAN family are indicative of a common head-to-head dimerization mode (Fig. 1). The various N- and C-terminal extensions of DAN family antagonists may provide additional contacts with the ligands as in the case of the N-terminal segment of Noggin. The structure of the Noggin–BMP-7 complex, with its coaxial two-fold symmetry, might also serve as a framework for binding by the Twisted gastrulation homodimer and perhaps Chordin/SOG.

Among the three general types of BMP antagonist (cystine knot dimers, Chordin/SOG and BAMBI/nma), both the cystine knot antagonists and BAMBI/nma seem to have arisen by gene duplication, with the former from an ancestral ligand-like protein²⁰ and the latter from an ancestral receptor-like protein. Thus, the extra-cellular regulation of BMP activity is unique among signalling pathways in that gene duplications of both ligand and receptor genes may have led to the evolution of structurally related antagonists to downregulate the activity of their agonist homologues. The surprising structural homology between agonists and antagonists provides evidence of the ancient origin and evolution of BMP-related signalling pathways in an ancestral primitive metazoan. □

Methods

Complex formation and crystallization

Recombinant human BMP-7, expressed in Chinese hamster ovary cells⁴³, was obtained from Curis Inc. Human Noggin protein⁴⁴, expressed in *Escherichia coli*, was obtained from Regeneron Pharmaceuticals in Noggin buffer (50 mM sodium citrate (pH 6.8), 150 mM NaCl, 1 mM magnesium acetate, 20% glycerol and 0.02% CHAPS). BMP-7 and Noggin were mixed at an equimolar ratio in a high salt buffer (0.7 M NaCl, 20 mM Tris-HCl (pH 7.4) and 1.8% CHAPS) and the resulting complex, a single heterodimeric species, was concentrated by ultrafiltration and purified by size-exclusion chromatography on a Superdex S75 column (Amersham Pharmacia). We crystallized Noggin–BMP-7 at 10 mg ml⁻¹ from hanging drops in 0.6 M disodium, 1.0 M monopotassium phosphate (~pH 6.5). Large single crystals grew overnight at ambient temperature in space group P4₁2₁2 with cell dimensions $a = b = 99.83 \text{ \AA}$, $c = 150.33 \text{ \AA}$.

Structure determination

All data were collected at 100 K on beamline 9-2 of the Stanford Synchrotron Radiation Laboratory (SSRL) from crystals soaked for roughly 2 d in reservoir solution with 25% glycerol. Bromide ions for MAD phasing⁴⁸ were introduced by rapidly soaking (~30 s) crystals in the same buffer containing 1.0 M NaBr before freezing in liquid nitrogen. We processed diffraction data with HKL⁴⁵ and located two sites using SOLVE⁴⁶ and eight more using MLPHARE. The phases were improved by DM⁴⁷ and the resulting maps were of sufficient quality to place the free BMP-7 model¹⁹. Alternating cycles of model building with O and refinement by REFMAC⁴⁷ resulted in a complete model of the Noggin protein, excluding seven residues of the disordered polyglycine segment. The asymmetric unit contains one BMP-7 and one Noggin (solvent volume of 75%).

Noggin variant expression, refolding and purification

Human Noggin variants were constructed from the *E. coli* expression vector pRG301. Inclusion body protein was denatured in 6 M guanidine-HCl, 50 mM Tris-HCl (pH 8.5), 1 mM EDTA and 10 mM dithiothreitol (DTT), exchanged into 6 M urea, 50 mM sodium acetate (pH 4.5) and 1 mM DTT, rapidly diluted in refolding buffer (1.5 M guanidine-HCl, 50 mM Tris-HCl (pH 8.0), 1 mM EDTA, 2 mM reduced glutathione and 0.2 mM oxidized glutathione) and incubated for 3 d at 4 °C. After refolding, the protein solution was adjusted to pH 7.0, concentrated and dialysed against Noggin buffer to remove misfolded protein by centrifugation. We purified wild-type dimer by ion exchange chromatography on an SP-Sepharose column (Amersham Pharmacia).

In vitro binding assay

Folded, purified dimers were conjugated to polyethylene glycol (PEG)–biotin by reaction with a 50-fold molar excess of Biotin–PEG–CO₂–NHS, MW 3400 (Shearwater) in Noggin buffer for 2 h on ice and isolated from residual PEG–biotin polymer by size-exclusion chromatography. BMP-7 was iodinated by NaI and chloramine-T. Binding of PEG–biotinylated Noggin to BMP-7 was done in 96-well plates (0.2- μ m, Durapore) that were pre-wetted with binding buffer (25 mM Tris-HCl (pH 7.5), 100 mM NaCl, 0.1% bovine serum albumin and 0.01% CHAPS). We incubated BMP-7 tracer with PEG–biotinylated Noggin, streptavidin–agarose and dilutions of unlabelled BMP-7 competitor. After 90 min, the plates were vacuum-filtered, washed three times and dried, and individual filters punched out and counted.

Chick limb bud assays

All chick embryos (*Gallus gallus*) were manipulated in the right leg bud at Hamburger and Hamilton stage 28, the left leg bud was used as a control. We carried out analyses after 30 h of treatment except for *in situ* analyses, which were done after 14 h. Heparin acrylic beads (Sigma) were used for digit growth assays as described³⁷. For Alcian Green staining, embryos were fixed in 5% trichloroacetic acid overnight before being incubated in 0.1%

dye. For *in situ* hybridizations with *sox9* (a gift from T. Sharpe), embryos were fixed in 4% paraformaldehyde overnight and processed as described⁴⁹. For vital staining, leg buds were treated with a solution of neutral red³⁵. For the TUNEL assay, we fixed limb buds in 4% paraformaldehyde overnight and treated tissue sections (16 μ m) as described in the *In Situ* Cell Death Detection kit (Roche).

Received 21 May; accepted 15 October 2002; doi:10.1038/nature01245.

- Massague, J. TGF- β signal transduction. *Annu. Rev. Biochem.* **67**, 753–791 (1998).
- Hogan, B. L. Bone morphogenetic proteins: multifunctional regulators of vertebrate development. *Genes Dev.* **10**, 1580–1594 (1996).
- Smith, W. C. TGF β inhibitors. *Trends Genet.* **15**, 3–5 (1999).
- Urist, M. R. Bone: formation by autoinduction. *Science* **150**, 893–899 (1965).
- Smith, W. C. & Harland, R. M. Expression cloning of noggin, a new dorsalizing factor localized to the Spemann organizer in *Xenopus* embryos. *Cell* **70**, 829–840 (1992).
- Hsu, D. R., Economides, A. N., Wang, X., Eimon, P. M. & Harland, R. M. The *Xenopus* dorsalizing factor Gremlin identifies a novel family of secreted proteins that antagonize BMP activities. *Mol. Cell* **1**, 673–683 (1998).
- Piccolo, S., Sasai, Y., Lu, B. & De Robertis, E. M. Dorsal/ventral patterning in *Xenopus*: inhibition of ventral signals by direct binding of chordin to BMP-4. *Cell* **86**, 589–598 (1996).
- Zimmerman, L. B., De Jesus-Escobar, J. M. & Harland, R. M. The Spemann organizer signal noggin binds and inactivates bone morphogenetic protein 4. *Cell* **86**, 599–606 (1996).
- Brunet, L. J., McMahon, J. A., McMahon, A. P. & Harland, R. M. Noggin, cartilage morphogenesis, and joint formation in the mammalian skeleton. *Science* **280**, 1455–1457 (1998).
- Gong, Y. *et al.* Heterozygous mutations in the gene encoding noggin affect human joint morphogenesis. *Nature Genet.* **21**, 302–304 (1999).
- Dixon, M. E., Armstrong, P., Stevens, D. B. & Bamshad, M. Identical mutations in NOG can cause either tarsal/carpal coalition syndrome or proximal symphalangism. *Genet. Med.* **3**, 349–353 (2001).
- Takahashi, T. *et al.* Mutations of the NOG gene in individuals with proximal symphalangism and multiple synostosis syndrome. *Clin. Genet.* **60**, 447–451 (2001).
- Lim, D. A. *et al.* Noggin antagonizes BMP signalling to create a niche for adult neurogenesis. *Neuron* **28**, 713–726 (2000).
- Ogawa, K. *et al.* Induction of a noggin-like gene by ectopic DV interaction during planarian regeneration. *Dev. Biol.* **250**, 59–70 (2002).
- Economides, A. N., Stahl, N. E., & Harland, R. M. Modified noggin polypeptide and compositions. US Patent 6,075,007 (2000).
- Paine-Saunders, S., Viviano, B. L., Economides, A. N. & Saunders, S. Heparan sulfate proteoglycans retain Noggin at the cell surface: a potential mechanism for shaping bone morphogenetic protein gradients. *J. Biol. Chem.* **277**, 2089–2096 (2002).
- Groppe, J. *et al.* Biochemical and biophysical characterization of refolded *Drosophila* DPP, a homolog of bone morphogenetic proteins 2 and 4. *J. Biol. Chem.* **273**, 29052–29065 (1998).
- Dauter, Z., Dauter, M. & Rajashankar, K. R. Novel approach to phasing proteins: derivatization by short cryo-soaking with halides. *Acta Crystallogr. D* **56**, 232–237 (2000).
- Griffith, D. L., Keck, P. C., Sampath, T. K., Rueger, D. C. & Carlson, W. D. Three-dimensional structure of recombinant human osteogenic protein 1: structural paradigm for the transforming growth factor β superfamily. *Proc. Natl Acad. Sci. USA* **93**, 878–883 (1996).
- Vitt, U. A., Hsu, S. Y. & Hsueh, A. J. Evolution and classification of cystine knot-containing hormones and related extracellular signaling molecules. *Mol. Endocrinol.* **15**, 681–694 (2001).
- Kirsch, T., Nickel, J. & Sebald, W. BMP-2 antagonists emerge from alterations in the low-affinity binding epitope for receptor BMPR-II. *EMBO J.* **19**, 3314–3324 (2000).
- Kirsch, T., Sebald, W. & Dreyer, M. K. Crystal structure of the BMP-2–BRIA ectodomain complex. *Nature Struct. Biol.* **7**, 492–496 (2000).
- Stanley, E. *et al.* DAN is a secreted glycoprotein related to *Xenopus* cerberus. *Mech. Dev.* **77**, 173–184 (1998).
- Isaacs, N. W. Cystine knots. *Curr. Opin. Struct. Biol.* **5**, 391–395 (1995).
- Sun, P. D. & Davies, D. R. The cystine-knot growth-factor superfamily. *Annu. Rev. Biophys. Biomol. Struct.* **24**, 269–291 (1995).
- Hatta, T. *et al.* Identification of the ligand-binding site of the BMP type IA receptor for BMP-4. *Biopolymers* **55**, 399–406 (2000).
- Capdevila, J. & Johnson, R. L. Endogenous and ectopic expression of noggin suggests a conserved mechanism for regulation of BMP function during limb and somite patterning. *Dev. Biol.* **197**, 205–217 (1998).
- Merino, R. *et al.* Morphogenesis of digits in the avian limb is controlled by FGFs, TGF β s, and noggin through BMP signaling. *Dev. Biol.* **200**, 35–45 (1998).
- Capdevila, J., Tsukui, T., Rodriguez Esteban, C., Zapavigna, V. & Izpisua Belmonte, J. C. Control of vertebrate limb outgrowth by the proximal factor Meis2 and distal antagonism of BMPs by Gremlin. *Mol. Cell* **4**, 839–849 (1999).
- Merino, R. *et al.* The BMP antagonist Gremlin regulates outgrowth, chondrogenesis and programmed cell death in the developing limb. *Development* **126**, 5515–5522 (1999).
- Zuniga, A., Harniss, A. P., McMahon, A. P. & Zeller, R. Signal relay by BMP antagonism controls the SHH/FGF4 feedback loop in vertebrate limb buds. *Nature* **401**, 598–602 (1999).
- Iemura, S. *et al.* Direct binding of follistatin to a complex of bone-morphogenetic protein and its receptor inhibits ventral and epidermal cell fates in early *Xenopus* embryo. *Proc. Natl Acad. Sci. USA* **95**, 9337–9342 (1998).
- Francis-West, P. H., Parish, J., Lee, K. & Archer, C. W. BMP/GDF-signalling interactions during synovial joint development. *Cell Tissue Res.* **296**, 111–119 (1999).
- Duprez, D. *et al.* Overexpression of BMP-2 and BMP-4 alters the size and shape of developing skeletal elements in the chick limb. *Mech. Dev.* **57**, 145–157 (1996).
- Macias, D. *et al.* Role of BMP-2 and OP-1 (BMP-7) in programmed cell death and skeletogenesis during chick limb development. *Development* **124**, 1109–1117 (1997).
- Enomoto-Iwamoto, M. *et al.* Bone morphogenetic protein signaling is required for maintenance of differentiated phenotype, control of proliferation, and hypertrophy in chondrocytes. *J. Cell Biol.* **140**, 409–418 (1998).
- Yokouchi, Y. *et al.* BMP-2/–4 mediate programmed cell death in chicken limb buds. *Development* **122**, 3725–3734 (1996).

38. Zou, H. & Niswander, L. Requirement for BMP signaling in interdigital apoptosis and scale formation. *Science* **272**, 738–741 (1996).
39. Gañan, Y., Macías, D., Duterte-Coquillaud, M., Ros, M. A. & Hurlé, J. M. Role of TGFβs and BMPs as signals controlling the position of the digits and the areas of interdigital cell death in the developing chick limb autopod. *Development* **122**, 2349–2357 (1996).
40. Healy, C., Uwanogho, D. & Sharpe, P. T. Regulation and role of Sox9 in cartilage formation. *Dev. Dyn.* **215**, 69–78 (1999).
41. Marcelino, J. *et al.* Human disease-causing NOG missense mutations: effects on noggin secretion, dimer formation, and bone morphogenetic protein binding. *Proc. Natl Acad. Sci. USA* **98**, 11353–11358 (2001).
42. Mangino, M., Flex, E., Digilio, M. C., Giannotti, A. & Dallapiccola, B. Identification of a novel NOG gene mutation (P35S) in an Italian family with symphalangism. *Hum. Mutat.* **19**, 308 (2002).
43. Sampath, T. K. *et al.* Recombinant human osteogenic protein-1 (hOP-1) induces new bone formation in vivo with a specific activity comparable with natural bovine osteogenic protein and stimulates osteoblast proliferation and differentiation in vitro. *J. Biol. Chem.* **267**, 20352–20362 (1992).
44. Valenzuela, D. M. *et al.* Identification of mammalian noggin and its expression in the adult nervous system. *J. Neurosci.* **15**, 6077–6084 (1995).
45. Otwinowski, Z. & Minor, W. Processing X-ray diffraction data collected in oscillation mode. *Methods Enzymol.* **276**, 307–326 (1997).
46. Terwilliger, T. C. & Berendzen, J. Automated MAD and MIR structure solution. *Acta Crystallogr. D* **55**, 849–861 (1999).
47. Collaborative Computational Project Number 4 The CCP4 suite: programs for protein crystallography. *Acta Crystallogr. D* **50**, 760–763 (1994).
48. Thompson, J. D., Higgins, D. G. & Gibson, T. J. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res.* **22**, 4673–4680 (1994).
49. Russell, R. B. & Barton, G. J. Multiple protein sequence alignment from tertiary structure comparison: assignment of global and residue confidence levels. *Proteins Struct. Funct. Genet.* **14**, 309–323 (1992).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank Curis for BMP-7; Regeneron Pharmaceuticals for Noggin; K. Baban for technical assistance; M. Austin for aid with crystal/cryo-optimization; G. Louie for phasing advice; and the staff at SSRL for help with data collection. The SSRL Structural Molecular Biology Program is supported by the Department of Energy and the NIH. J. Groppe is grateful to K. Kirschner and T. Kiefhaber for their initial support. J. Greenwald acknowledges support from a National Cancer Institute Training Grant. This work was supported by grants from the Swiss National Science Foundation and the Kantons Basel (M.A.), BioCell, National Science Foundation, the G. Harold and Leila Y. Mathers Charitable Foundation, and the Fundacao Calouste Gulbenkian and Fundacao para Ciencia e Tecnologia (J.R.L., J.C.I.B.), and the National Institutes of Health (S.C., W.W.V.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C. (e-mail: choe@salk.edu). Coordinates for the structure have been deposited in the Protein Data Bank (accession code 1M4U).

Formation of Kuiper-belt binaries by dynamical friction and three-body encounters

Peter Goldreich^{*†}, Yoram Lithwick^{†‡} & Re'em Sari[†]

^{*} School of Natural Science, Institute for Advanced Study, Princeton, New Jersey 08540, USA

[†] Theoretical Astrophysics, Caltech 130-33, Pasadena, California 91125, USA

[‡] Astronomy Department, University of California at Berkeley, Berkeley, California 94720, USA

The Kuiper belt is a disk of icy bodies that orbit the Sun beyond Neptune¹; the largest known members are Pluto and its companion Charon. A few per cent of Kuiper-belt bodies have recently been found to be binaries with wide separations and mass ratios of the order of unity^{2–8}. Collisions were too infrequent to account for the observed number of binaries⁹, implying that these binaries formed through collisionless interactions mediated by gravity. These interactions are likely to have been most effective during the period of runaway accretion, early in the Solar System's history. Here we show that a transient binary forms when two large bodies penetrate one another's Hill sphere (the region where their mutual forces are larger than the tidal force of the Sun). The loss of energy needed to stabilize the binary orbit can then occur either through dynamical friction from surrounding small bodies, or through the gravitational scattering of a third large body. Our estimates slightly favour the former mechanism. We predict that five per cent of Kuiper-belt objects are binaries with apparent separations greater than 0.2 arcsec, and that most are in tighter binaries or systems of higher multiplicity.

We first outline a simple model for runaway accretion. Accretion in the Kuiper belt did not reach completion, but was terminated when the velocity dispersion of its members was increased by an as-yet undetermined process¹⁰. Bodies with a wide range of sizes were created during runaway accretion because the logarithmic mass growth rate, $M^{-1} dM/dt$, was an increasing function of M .

Our main simplification is to consider only two groups of bodies: small ones, containing most of the total mass, and large ones, contributing a small fraction of it. The latter group is identified with the currently observed population of Kuiper-belt objects. Dynamical processes acting during runaway accretion include: viscous stirring of all bodies by the large ones; dynamical friction by the small bodies which damps the random motions of the large ones; and physical collisions, which lead to accretion and fragmentation.

The Methods section describes the evolution of the radii, R , of the large bodies, and the velocity dispersions, v and u , of the large and small bodies, respectively. Two dimensionless parameters determine the appropriate regime of runaway accretion: $\theta_\odot \approx 10^{-4}$, the angle subtended by the solar radius as seen from the Kuiper belt, and Σ/σ , the ratio of the surface densities of the large to small bodies. On the basis of Kuiper-belt surveys^{11–13} that detect only large bodies, we estimate $R \approx 100$ km and $\Sigma \approx 3 \times 10^{-4} \text{ g cm}^{-2}$. Extrapolating the minimum-mass solar nebula surface density to 40 AU yields $\sigma \approx 0.3 \text{ g cm}^{-2}$. The resulting ratio of $\Sigma/\sigma \approx 10^{-3}$ is compatible with Kuiper-belt simulations^{14–16}. Our investigation is simplified by using the same density ($\rho \approx 1 \text{ g cm}^{-3}$) for both the Sun and the Kuiper-belt objects.

Large bodies grow primarily by accreting small ones; small bodies do not grow appreciably on the growth timescale of R . Small bodies are viscously stirred by large ones, with their velocity u growing on the same timescale as R grows. Mutual viscous stirring increases the velocity v of large bodies, but dynamical friction from the sea of small bodies maintains $v < u$. Both stirring and damping rates of v

are faster than the growth rate of R , so v is in quasi-steady state. Quantitatively, we find

$$\frac{v}{v_H} \approx \frac{1}{\theta_\odot} \left(\frac{\Sigma}{\sigma} \right)^{3/2} \approx 0.3 \text{ and } \frac{u}{v_H} \approx \left(\frac{1}{\theta_\odot} \frac{\Sigma}{\sigma} \right)^{1/2} \approx 3 \quad (1)$$

where v_H is the Hill velocity of the large bodies: $v_H \approx \Omega_\odot R_H$, where $\Omega_\odot \approx 1/40 \text{ yr}$ is the orbital frequency around the Sun, and $R_H \approx \theta_\odot^{-1} R \approx 10^4 R$ is the Hill radius. These results are in accord with the most recent simulations¹⁶.

We identify two distinct channels for binary formation. Because $v < v_H$, both begin when two large bodies penetrate one another's Hill sphere. This occurs at a rate

$$\frac{\Sigma}{M} R_H^2 \Omega_\odot \approx \frac{\Sigma}{\rho R} \left(\frac{1}{\theta_\odot} \right)^2 \Omega_\odot \approx 10^{-4} \text{ yr}^{-1} \quad (2)$$

per large body. Stabilization of such a transient binary requires energy loss on timescale $\Omega_\odot^{-1}$. This can be achieved either through dynamical friction from small bodies, a mechanism which we term L^2 s (for two large bodies and the sea of small bodies), or by interaction with a third large body, which we denote L^3 (for three large bodies).

(1) The L^2 s channel. Dynamical friction from small bodies during time $\Omega_\odot^{-1}$ results in a fractional energy loss:

$$\frac{\Delta E}{M v_H^2} \approx \frac{\sigma}{\rho R} \left(\frac{\sigma}{\Sigma} \right)^2 \approx 0.03 \quad (3)$$

This is also the fraction of transient binaries that become bound. Numerical simulations designed to test this assertion are shown in Fig. 1. We integrate the equations of motion including dynamical friction under the Hill approximation in a frame rotating at the average mean motion. It is evident that capture is only possible within distinct ranges of impact parameters, whose widths are roughly proportional to the drag. More precisely, the symbol C denotes the ratio of the linear measure of the impact parameters that result in capture to $3^{1/3} R_H$, and the symbol D denotes the fractional decrease in velocity due to drag over a time $\Omega_\odot^{-1}$, essentially half the expression in equation (3). Clearly $C \approx D$ to better than a factor of two, verifying our assertion. A sample of similar calculations starting from orbits with finite velocity $v < v_H$ gives similar results. Thus, this mechanism leads to a binary formation rate per large body:

$$FR_{L^2s} \approx \left(\frac{\sigma}{\rho R} \right)^2 \left(\frac{\sigma}{\Sigma} \right)^2 \left(\frac{1}{\theta_\odot} \right)^2 \Omega_\odot \approx 3 \times 10^{-6} \text{ yr}^{-1} \quad (4)$$

(2) The L^3 channel. The probability that a third large body joins the Hill sphere of a transient binary during its lifetime is:

$$\frac{\Sigma}{\rho R} \left(\frac{1}{\theta_\odot} \right)^2 \approx 3 \times 10^{-3} \quad (5)$$

A significant fraction of these triplets will result in a bound binary. Therefore, the binary formation rate per large body via the L^3 channel is:

$$FR_{L^3} \approx \left(\frac{\Sigma}{\rho R} \right)^2 \left(\frac{1}{\theta_\odot} \right)^4 \Omega_\odot \approx 3 \times 10^{-7} \text{ yr}^{-1} \quad (6)$$

With our parameters, the ratio

$$\frac{FR_{L^3}}{FR_{L^2s}} \approx \left(\frac{\Sigma}{\sigma} \right)^3 \left(\frac{1}{\theta_\odot} \right)^2 \approx 0.1 \quad (7)$$

is smaller than unity. Thus from here on we take $FR = FR_{L^2s}$; for smaller σ/Σ , $FR_{L^3} > FR_{L^2s}$, and that case can be treated in a similar manner. The ratio of FR to $R^{-1} dR/dt$ is

$$\frac{\sigma}{\rho R} \left(\frac{1}{\theta_\odot} \right)^2 \approx 3 \quad (8)$$

so binaries form at a rate comparable to the large bodies' growth rate.

An important outcome of our binary formation model is that it allows us to predict the fraction of Kuiper-belt objects in binaries, as well as the distribution of their semimajor axes. The semimajor axis, a , shrinks at a rate $a^{-1}|da/dt|$ owing to dynamical friction. Let $p(a)$ be the probability density of finding a large body in a binary with semimajor axis a . Then,

$$a p(a) \approx \frac{FR}{a^{-1}|da/dt|} \quad (9)$$

In the region where the small bodies are unaffected by the large one, $r_u < a < R_H$ (see equation (23)), $a^{-1}|da/dt|$ is given by equation (21). Thus the probability per logarithmic band in a is independent of a :

$$a p(a) \approx \frac{\Sigma}{\rho R} \left(\frac{1}{\theta_\odot} \right)^2 \approx 3 \times 10^{-3} \quad (10)$$

For $R < a < r_u$, both the number density and the velocity dispersion of small bodies are enhanced by a factor $(r_u/a)^{1/2}$; see equations (24) and (25). Therefore, $a^{-1}|da/dt|$ is reduced by a factor a/r_u in comparison to its constant value for $r_u < a < R_H$

given by equation (21). Thus in this interval of semimajor axis:

$$a p(a) \approx \frac{\sigma}{\rho R} \left(\frac{1}{\theta_\odot} \right)^2 \frac{R}{a} \approx 3 \left(\frac{R}{a} \right) \quad (11)$$

The timescale for a binary to spiral in until contact is given by $a(da/dt)^{-1}$ evaluated at $a = R$ which equals the growth timescale for an isolated body $R(dR/dt)^{-1}$. Contact occurs on the same timescale as that during which large bodies grow by accreting small ones. We define a critical separation

$$a_{\text{crit}} \approx \frac{\sigma}{\rho \theta_\odot^2} \approx 300 \text{ km} \quad (12)$$

inside of which $a p(a) > 1$. For $R < a_{\text{crit}}$, accretion by binary inspiral could make a substantial contribution to the growth of large bodies and systems of higher multiplicity might be common.

Exchange reactions, in which the lighter member of a binary is replaced by a heavier body that passes through the system, are rare occurrences. Large bodies pass through an existing binary of semimajor axis R_H at a rate given by equation (2). This is smaller than the orbital decay rate by the factor $(\Sigma/\sigma)^3 \theta_\odot^{-2} \approx 0.1$. For

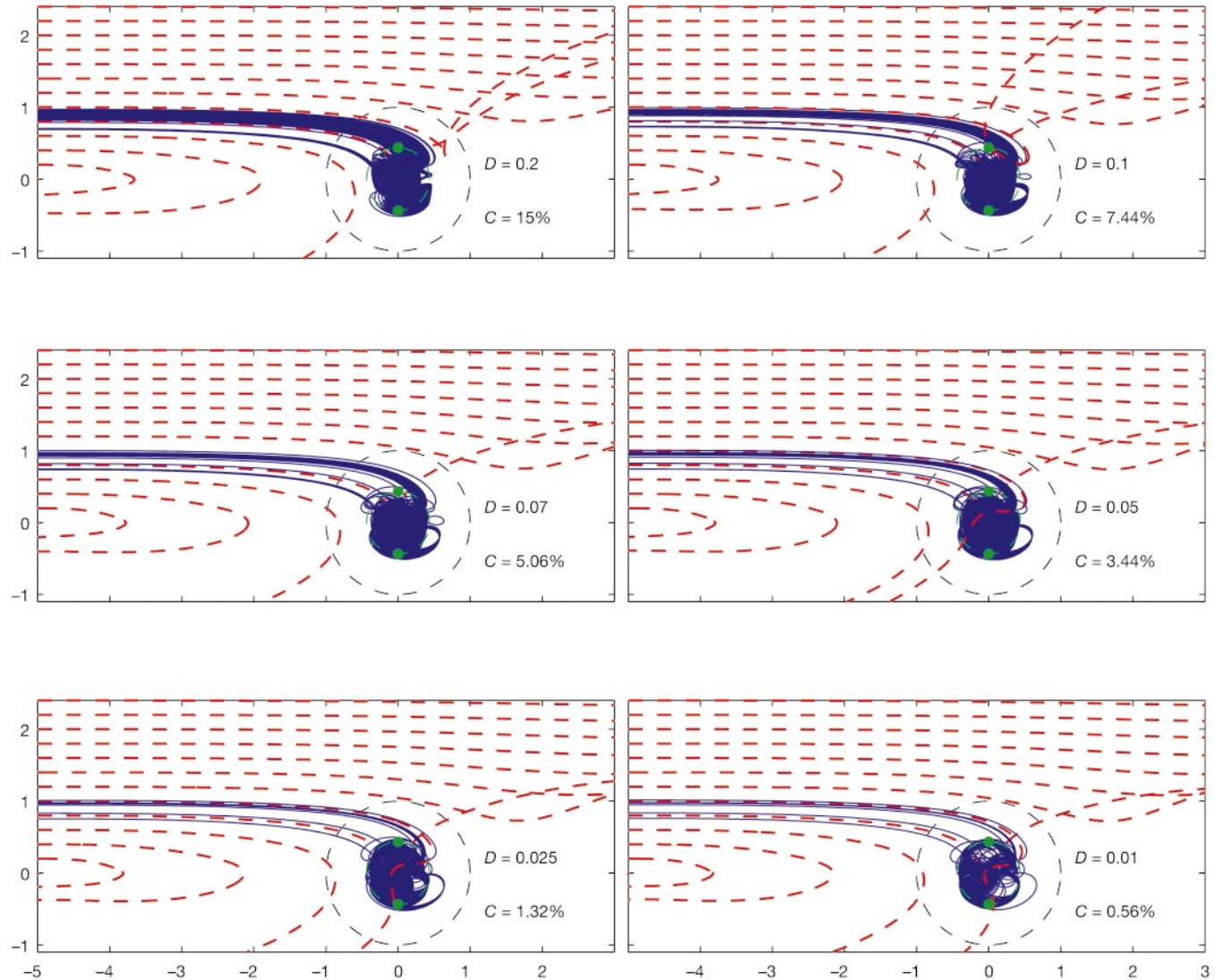


Figure 1 Numerical simulations of binary formation by dynamical friction for six values of the drag coefficient D . Two equal-mass bodies approach each other on initially circular orbits. The centre of mass is fixed at $(0,0)$. We plot only the trajectories of one of the bodies; the trajectories of the other body are related by reflection through the origin. Dashed red trajectories did not form binaries, and solid blue ones did. The axes are in

units of $3^{1/3}R_H$, where R_H is the Hill radius. The Lagrange points of the corresponding restricted three-body problem are marked by green dots and the Hill sphere by a dashed green circle. Every capture orbit is plotted, but only a sample of the non-capture orbits are shown. C measures the capture probability.

smaller semimajor axes, exchange reactions are rarer still. Near coalescence, only $(\Sigma/\sigma)^{1/2} \approx 0.03$ large bodies pass through the binary during an orbital decay time.

In summary, we propose that the observed Kuiper-belt binaries—which are widely spaced and have mass ratios of the order of unity—formed during runaway accretion. A fraction of the large bodies that entered one another's Hill sphere became bound as the result of energy lost to small bodies by dynamical friction (L^2 s channel) or by interaction with a third large body (L^3 channel). The timescale for a large body to become bound to a similar companion by the dominant L^2 s process was:

$$\left(\frac{\rho R}{\sigma}\right)^2 \left(\frac{\Sigma}{\sigma}\right) \frac{\theta_{\odot}^2}{\Omega_{\odot}} \approx 3 \times 10^5 \text{ yr} \quad (13)$$

Further dynamical friction hardened the binaries. The timescale to achieve contact was that during which isolated large bodies grew:

$$\frac{\rho R}{\sigma} \left(\frac{\Sigma}{\sigma}\right) \frac{1}{\Omega_{\odot}} \approx 10^6 \text{ yr} \quad (14)$$

Initially the inspiral was at a constant rate, but it slowed down inside r_u where the small bodies' number density and velocity dispersion were enhanced above their background values. We deduce that the probability that a large body is part of a binary with separation greater than r_u , about $3''$ of angular separation, is:

$$\frac{\Sigma}{\rho R} \left(\frac{1}{\theta_{\odot}}\right)^2 \approx 3 \times 10^{-3} \quad (15)$$

Inward of r_u , the binary probability per logarithmic interval of semimajor axis increases inversely with semimajor axis. Because $a p(a)$, as given by equation (11), exceeds unity for $R < a < a_{\text{crit}} \approx 300 \text{ km}$, systems with higher multiplicities would exist if such are stable. For the resolution of the Hubble Space Telescope (HST) survey ($\sim 0.2''$), we predict a binary fraction of about 5%, roughly compatible with observations (M.E. Brown, personal communication). Our prediction that close binaries are common could be tested by monitoring the brightness of Kuiper-belt objects for evidence of eclipses and/or fast rotation. The latter is a consequence, but not a unique signature, of binary mergers.

Weidenschilling has proposed¹⁷ a model for the formation of Kuiper-belt binaries that is different from either of our two mechanisms L^2 s and L^3 . He considers the physical collision of two large bodies inside the Hill sphere of a third large body. To account for the observed binary frequency, he is forced to assume that, at the time of binary formation, the number of large bodies exceeded its current value by two orders of magnitude. No explanation is offered for the disappearance of most of the large bodies since that time. In fact, simulations of the formation of Kuiper-belt objects¹⁶ show that more than 99% of the total mass was in much smaller bodies. Moreover, even if the Kuiper belt was populated by 100 times more large bodies than are at present there, the binary formation rate from Weidenschilling's model would be negligible relative to that from collisionless gravitational interactions between three bodies, as evaluated in our equation (6).

By contrast, our collisionless mechanism for binary formation is adequate to account for the formation of the observed fraction of Kuiper-belt binaries using no more than the observationally determined number of large Kuiper-belt objects. The surface density of small bodies required to provide dynamical friction is just that required for the large Kuiper-belt objects to have grown by accretion. Finally, Weidenschilling predicts that the number of binaries should increase with increasing semimajor axis, whereas we predict the opposite. These predictions will be subject to observational tests in the near future.

Our numerical results are crude, as we have discarded numerical coefficients and logarithmic factors. Therefore, more credence should be accorded to functional dependences than to precise

numerical predictions. We foresee numerous extensions to the present investigation. In our analysis we estimate the surface density ratio of large to small bodies from the observational census of Kuiper-belt objects together with an extrapolation of the surface density of the minimum-mass solar nebula. It would be an improvement to have a theoretical understanding of how this ratio evolves during runaway accretion. Also, we elaborated the binary evolution model with a given set of parameters, specifically $\theta_{\odot}$ and Σ/σ . Different scalings that apply for other parameter regimes remain to be worked out.

Additional predictions suitable for testing by future observations of Kuiper-belt binaries could be made by extending our model in a variety of ways. Relaxing the restriction to two mass groups would allow us to investigate the mass ratio distribution of binary components and its dependence on semimajor axis. We predict the formation of multiple systems for bodies with $R < a_{\text{crit}}$. Numerical experiments could establish what types of systems survive disruption due to mutual interactions of their members. We could also predict eccentricity and inclination distributions. This would require a more detailed look at the anisotropy of the velocity distribution of small bodies, especially for radii $r < r_u$. Anisotropy leads to tensorial dynamical friction in which the force is not antiparallel to the velocity¹⁸. We can already say that the eccentricities are likely to be large. They start large following binary capture, and because the magnitude of the dynamical friction force increases outwards for radii $r < r_u$, they should remain large in this region.

Numerical simulations of the Hill problem that include both large and small bodies would help determine the capture probability due to dynamical friction. Computations involving three massive bodies are needed for evaluation of capture probabilities by that channel. These have been computed for stars in globular clusters¹⁹, but in our regime the Hill sphere must be taken into account.

The two largest known Kuiper-belt objects, Pluto and Charon, form a binary. Collision coupled with tidal evolution is the generally accepted mode of formation for this system. Our model for binary formation provides another candidate. It may also apply to the formation of some binaries in the asteroid belt. $\square$

Methods

Runaway accretion

Gravitational interactions determine the accretion rates and velocity dispersions of both large and small bodies. Simplified equations, abstracted from more complete treatments in the literature^{20,21}, read:

$$\frac{1}{R} \frac{dR}{dt} \approx \frac{\Sigma \Omega_{\odot}}{\rho R} F_{M-M} + \frac{\sigma \Omega_{\odot}}{\rho R} F_{M-m}, \quad (16)$$

$$\frac{1}{s} \frac{ds}{dt} \approx \frac{\sigma \Omega_{\odot}}{\rho s}, \quad (17)$$

$$\frac{1}{v} \frac{dv}{dt} \approx \frac{\Sigma \Omega_{\odot}}{\rho R} F_{M-M}^2 - \frac{\sigma \Omega_{\odot}}{\rho R} F_{M-m}^2, \quad (18)$$

$$\frac{1}{u} \frac{du}{dt} \approx \frac{\Sigma \Omega_{\odot}}{\rho R} F_{M-m}^2 - \frac{\sigma \Omega_{\odot}}{\rho s}, \quad (19)$$

where s denotes the radii of small bodies. To simplify matters, we use a single velocity dispersion in place of three that are needed to characterize a triaxial velocity ellipsoid.

The expressions $\Sigma \Omega_{\odot}/\rho R$ and $\sigma \Omega_{\odot}/\rho s$ give the kinematic (ignoring gravitational focusing) collision rates of large bodies onto large bodies and small bodies onto small bodies; $\sigma \Omega_{\odot}/\rho R$ is the kinematic collision rate of small bodies onto a large one multiplied by the mass ratio $(s/R)^3$. The factors F_{M-M} and F_{M-m} in equation (16) account for the enhancements of the physical collision rates owing to gravitational focusing by large bodies of the trajectories of incoming large and small bodies, respectively:

$$F_{M-m} \approx \begin{cases} 1 & \theta_{\odot}^{-1/2} v_H < u \\ \theta_{\odot}^{-1} (v_H/u)^2 & v_H < u < \theta_{\odot}^{-1/2} v_H \\ \theta_{\odot}^{-1} (v_H/u) & \theta_{\odot}^{1/2} v_H < u < v_H \end{cases} \quad (20)$$

The expression $\theta_{\odot}^{-1/2} v_H$ is the escape velocity. F_{M-M} is obtained by replacing u with v . Cross-sections for large deflections exceed those for physical collisions when gravitational focusing is significant. The factors F_{M-M}^2 and F_{M-m}^2 that appear in equations (18) and (19) are the ratios of those cross-sections to the geometrical ones.

In equation (16), the two terms on the right-hand side account for the radius growth of

the large bodies by the accretion of large and small bodies. The single term on the right-hand side of equation (17) describes how the radii of small bodies increase as the result of coalescence under conditions of negligible gravitational focusing. In equation (18), the first term describes the viscous stirring of large bodies due to mutual gravitational deflections, whereas the second term accounts for the damping of the large bodies' velocity dispersion by dynamical friction resulting from their gravitational interactions with the small bodies. Equation (19) shows how the velocity dispersion of the small bodies evolves under viscous stirring by large bodies and damping due to collisions between small bodies.

We assume that $(\Sigma/\sigma)^{2/3} < \theta_0 < \Sigma/\sigma < 1$. In this regime of runaway accretion, the rates of gravitational stirring and dynamical friction acting on the large bodies are much greater than the large bodies' growth rates. Moreover, provided the radii of the small bodies s satisfy $(\Sigma/\sigma)R < s < (\Sigma/\sigma)^{1/3}R$, or $0.1 \text{ km} < s < 10 \text{ km}$, their growth rates are negligible compared with those of the large bodies, and collisional damping of their velocities is unimportant on the timescale of viscous stirring. Under these approximations, the solution to equations (16)–(19) is given in equation (1). In this solution, the balanced rates of viscous stirring and dynamical friction for large bodies satisfy:

$$-\frac{1}{v} \frac{dv}{dt} \bigg|_{df} = \frac{1}{v} \frac{dv}{dt} \bigg|_{vs} \approx \frac{\sigma}{\rho R} \left(\frac{\sigma}{\Sigma} \right)^2 \Omega_0 \quad (21)$$

Large bodies grow predominantly by accreting small bodies; the small bodies' velocity dispersion evolves mainly by viscous stirring provided by the large bodies with negligible collisional losses. These two processes proceed at equal rates:

$$\frac{1}{R} \frac{dR}{dt} \approx \frac{1}{u} \frac{du}{dt} \approx \frac{\sigma}{\rho R} \left(\frac{\sigma}{\Sigma} \right) \Omega_0 \quad (22)$$

Spatial distribution of small bodies

Small bodies that are on hyperbolic orbits with respect to a large body are affected at separations smaller than

$$r_u \approx GM/u^2 \approx (\sigma/\Sigma)R < R_H \quad (23)$$

For separations $r < r_u$, the velocity dispersion of the small bodies is

$$(GM/r)^{1/2} \approx (r_u/r)^{1/2} u \quad (24)$$

In addition, because the impact parameter for a small body to arrive at radius r is $b \approx (r_u r)^{1/2}$, the continuity equation implies that for $r < r_u$ the small bodies' number density is enhanced by a factor:

$$\frac{b^2 u}{r^2 u (r_u/r)^{1/2}} \approx \left(\frac{r_u}{r} \right)^{1/2} \quad (25)$$

Received 29 August; accepted 17 October 2002; doi:10.1038/nature01227.

- Luu, J. X. & Jewitt, D. C. Trans-neptunian objects: Relics from the accretion disk of the sun. *Annu. Rev. Astron. Astrophys.* **40**, 63–101 (2002).
- Elliot, J. L., Kern, S. D., Osip, D. J. & Burles, S. M. 2001 QT₂₉₇. *IAU Circ. No. 7733* (2001).
- Kavelaars, J. J., Petit, J.-M., Gladman, B. & Holman, M. 2001 QW₃₂₂. *IAU Circ. No. 7749* (2001).
- Brown, M. E. & Trujillo, C. A. (26308) 1998 SM₁₆₅. *IAU Circ. No. 7807* (2002).
- Trujillo, C. A. & Brown, M. E. 1999 TC₃₆. *IAU Circ. No. 7787* (2002).
- Noll, K. et al. 2000 CF₁₀₅. *IAU Circ. No. 7857* (2002).
- Noll, K. et al. 1997 CQ₂₉. *IAU Circ. No. 7824* (2002).
- Veillet, C. et al. The binary Kuiper-belt object 1998 WW₃₁. *Nature* **416**, 711–713 (2002).
- Stern, S. A. Implications regarding the energetics of the collisional formation of Kuiper belt satellites. Preprint astro-ph/0206104 at xxx.lanl.gov (2002).
- Morbidelli, A., Jacob, C. & Petit, J.-M. Planetary embryos never formed in the Kuiper belt. *Icarus* **157**, 241–248 (2002).
- Trujillo, C. A., Jewitt, D. C. & Luu, J. X. Properties of the trans-neptunian belt: Statistics from the Canada-France-Hawaii telescope survey. *Astron. J.* **122**, 457–473 (2001).
- Gladman, B. et al. Pencil-beam surveys for faint trans-neptunian objects. *Astron. J.* **116**, 2042–2054 (1998).
- Chiang, E. I. & Brown, M. E. Keck pencil-beam survey for faint Kuiper belt objects. *Astron. J.* **118**, 1411–1422 (1999).
- Kenyon, S. J. & Luu, J. X. Accretion in the early Kuiper belt. I. Coagulation and velocity evolution. *Astron. J.* **115**, 2136–2160 (1998).
- Kenyon, S. J. & Luu, J. X. Accretion in the early Kuiper belt. II. Fragmentation. *Astron. J.* **118**, 1101–1119 (1999).
- Kenyon, S. J. Planet formation in the outer solar system. *Publ. Astron. Soc. Pacif.* **114**, 265–283 (2002).
- Weidenschilling, S. J. On the origin of binary transneptunian objects. *Icarus* **160**, 212–215 (2002).
- Binney, J. Dynamical friction in aspherical clusters. *Mon. Not. R. Astron. Soc.* **181**, 735–746 (1977).
- Hut, P. & Bahcall, J. N. Binary-single star scattering. I—Numerical experiments for equal masses. *Astrophys. J.* **268**, 319–341 (1983).
- Wetherill, G. W. & Stewart, G. R. Accumulation of a swarm of small planetesimals. *Icarus* **77**, 330–357 (1989).
- Wetherill, G. W. & Stewart, G. R. Formation of planetary embryos—Effects of fragmentation, low relative velocity, and independent variation of eccentricity and inclination. *Icarus* **106**, 190–209 (1993).

Acknowledgements We thank S. Kenyon for discussions about the dynamics of accretion. This work was supported by the NSF and NASA. R.S. is supported by the Fairchild Foundation.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.S. (e-mail: sari@tapir.caltech.edu).

Zero-resistance states induced by electromagnetic-wave excitation in GaAs/AlGaAs heterostructures

Ramesh G. Mani^{*†}, Jürgen H. Smet[‡], Klaus von Klitzing[‡], Venkatesh Narayanamurti^{*‡}, William B. Johnson[§] & Vladimir Umansky^{||}

^{*} Gordon McKay Laboratory of Applied Science, Harvard University, 9 Oxford Street, Cambridge, Massachusetts 02138, USA

[†] Max-Planck-Institut für Festkörperforschung, Heisenbergstrasse 1, 70569 Stuttgart, Germany

[‡] Pierce Hall, Harvard University, 29 Oxford Street, Cambridge, Massachusetts 02138, USA

[§] Laboratory for Physical Sciences, University of Maryland, College Park, Maryland 20740, USA

^{||} Braun Center for Submicron Research, Weizmann Institute, Rehovot 76100, Israel

The observation of vanishing electrical resistance in condensed matter has led to the discovery of new phenomena such as, for example, superconductivity, where a zero-resistance state can be detected in a metal below a transition temperature T_c (ref. 1). More recently, quantum Hall effects were discovered from investigations of zero-resistance states at low temperatures and high magnetic fields in two-dimensional electron systems (2DESs)^{2–4}. In quantum Hall systems and superconductors, zero-resistance states often coincide with the appearance of a gap in the energy spectrum^{1,2,4}. Here we report the observation of zero-resistance states and energy gaps in a surprising setting⁵: ultrahigh-mobility GaAs/AlGaAs heterostructures that contain a 2DES exhibit vanishing diagonal resistance without Hall resistance quantization at low temperatures and low magnetic fields when the specimen is subjected to electromagnetic wave excitation. Zero-resistance states occur about magnetic fields $B = 4/5B_f$ and $B = 4/9B_f$, where $B_f = 2\pi f m^*/e$, m^* is the electron mass, e is the electron charge, and f is the electromagnetic-wave frequency. Activated transport measurements on the resistance minima also indicate an energy gap at the Fermi level⁶. The results suggest an unexpected radiation-induced, electronic-state-transition in the GaAs/AlGaAs 2DES.

Hall bars of width w , with $50 \leq w \leq 200 \mu\text{m}$, and square-shaped specimens up to $\sim 3 \times 3 \text{ mm}^2$ were fabricated from high quality GaAs/AlGaAs heterostructures exhibiting an electron density $n(4.2\text{K}) \approx 3 \times 10^{11} \text{ cm}^{-2}$ and a mobility $\mu(1.5\text{K}) \approx 1.5 \times 10^7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. Typically, a specimen including either Au/Ge-Ni or In contacts was mounted inside a waveguide, immersed in pumped liquid helium, and irradiated with electromagnetic waves (microwaves) over the frequency range $27 \leq f \leq 115 \text{ GHz}$, at an estimated power level of $\leq 1 \text{ mW}$, over a cross-sectional area of $\leq 135 \text{ mm}^2$ in the vicinity of the sample. The range of f was spanned piecewise using an array of sources. The current axis in the specimen was oriented either parallel or perpendicular to the electromagnetic-wave polarization axis of the waveguide. Observed experimental features appeared to be insensitive to the type of contacts, the shape of the sample, the current orientation with respect to the polarization axis, and the magnitude of the current. Conversely, these physical phenomena were quite sensitive to the temperature, the microwave frequency, and the radiation power. The systematic dependence upon these latter variables served to rule out spurious effects.

Figure 1a shows the diagonal (R_{xx}) and Hall (R_{xy}) resistances, which were measured in the four-terminal configuration, using standard low-frequency a.c. lock-in techniques. Notably, R_{xx} and R_{xy} exhibit the usual quantum Hall behaviour for $B \geq 0.4 \text{ T}$ at $f =$

103.5 GHz (refs 2–4). In contrast, at $B < 0.4$ T (Fig. 1a inset), a radiation-induced signal occurs and, remarkably, the resistance vanishes over a broad interval of B around $B = 0.198$ T. Here, the scale factor to convert R_{xx} to the resistivity ρ_{xx} is 1.

Further high-resolution measurements are shown in Fig. 1b. Without electromagnetic excitation, R_{xx} exhibits Shubnikov–

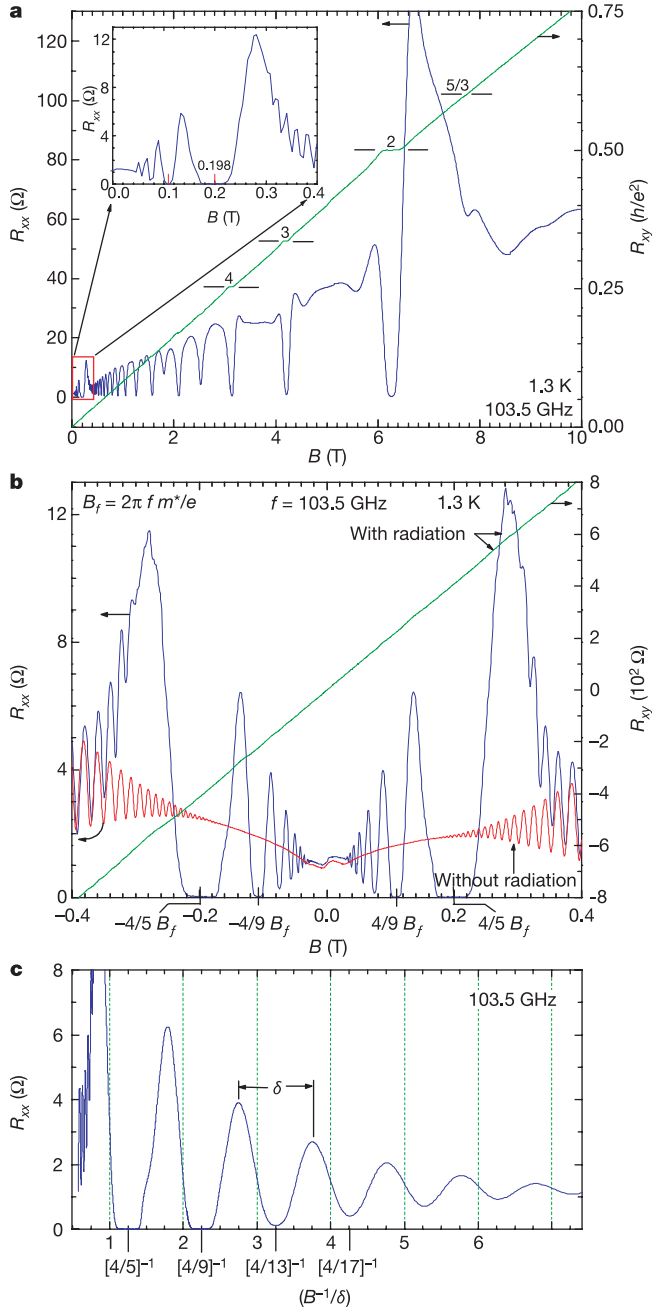


Figure 1 The Hall effect and the magnetoresistance in a high-mobility 2DES, with and without electromagnetic-wave excitation. **a**, The Hall (R_{xy}) and diagonal (R_{xx}) resistances in a GaAs/AlGaAs heterostructure under excitation at 103.5 GHz. Quantum Hall effects occur at high B as $R_{xx} \rightarrow 0$. Inset, an expanded view of the low- B data showing $R_{xx} \rightarrow 0$ in the vicinity of 0.198 T. **b**, Data over low magnetic fields obtained both with (w/) and without (w/o) radiation at 103.5 GHz. Note the vanishing resistance under radiation in the vicinity of $\pm 4/5 B_f$ and $\pm 4/9 B_f$, and the linear Hall effect over corresponding B intervals. In contrast, quantum Hall systems typically exhibit plateaus in the Hall resistance over the B intervals where vanishing resistance is observed. **c**, A normalized B^{-1} plot of the low- B data. Resistance minima occur at $[4/(4j+1)]^{-1}$ with $j = 1, 2, 3, \dots$. Here, δ is the oscillatory period in B^{-1} , which agrees with B_f^{-1} within experimental uncertainties ($\sim 1\%$). Note that the $j = 1$ and $j = 2$ states exhibit vanishing resistance.

deHaas oscillations for $|B| \geq 0.2$ T. Radiation induces additional oscillations, which might be attributed to scattering between Landau levels⁷, as in the magneto-phonon effect⁸. Yet, the data reveal that, at the minima, R_{xx} falls well below the resistance measured without radiation and it vanishes around magnetic fields of $4/5 B_f$ and $4/9 B_f$ set by $B_f = 2\pi f m^*/e$ with $m^* = 0.067m$, the GaAs electron effective mass⁹. Although these zero-resistance states exhibit a flat bottom as in the quantum Hall regime², R_{xy} does not exhibit plateaus over the same B interval. A normalized B^{-1} plot (Fig. 1c) demonstrates periodicity in B^{-1} with minima that satisfy $B_{\min}^{-1}/\delta = [4/(4j+1)]^{-1}$ with $j = 1, 2, 3, \dots$. Here, δ is the oscillatory period in B^{-1} , which agrees with B_f^{-1} within experimental uncertainties ($\sim 1\%$). Notably, half-cycle plots also confirmed a '1/4-cycle phase shift' of the extrema with respect to integral B_f^{-1} , which is indicated by Fig. 1c.

Figure 2 shows the f dependence of the R_{xx} oscillations. For f up to 40 GHz (not shown), R_{xx} exhibited just a minimum at $4/5 B_f$. A data-fit with exponentially damped sinusoids¹⁰ suggested that the resistance-oscillation frequency F increases linearly with f , and that the oscillation amplitude decay in B^{-1} is characterized by an f -independent damping parameter. An evaluation of m^* using $dF/df = 2\pi m^*/e$ yielded $m^*/m = 0.067$, consistent with expectations for GaAs (refs 9, 11, 12). Over the intermediate f range (Fig. 2a), a zero-resistance state is first observed at $4/5 B_f$; this shifts to higher B with increasing f . Similar behaviour continues onto the

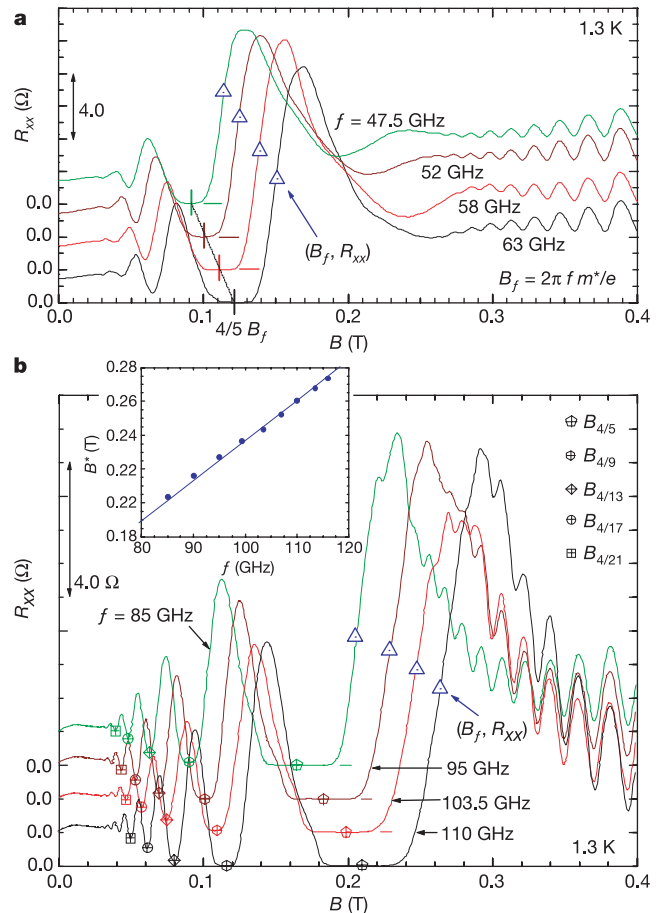


Figure 2 The development of the radiation-induced zero-resistance states with the electromagnetic-wave frequency, f . **a**, Over the intermediate f -range, a zero-resistance state occurs at $4/5 B_f$. The triangles mark R_{xx} at the magnetic field $B_f = 2\pi f m^*/e$, where the oscillatory curves exhibit neither a maximum nor a minimum. **b**, Over the highest f -range, a zero-resistance state is seen at both $4/5 B_f$ and $4/9 B_f$, marked here as $B_{4/5}$ and $B_{4/9}$. Inset, B^* calculated using the (weighted) first five resistance minima as a function of f . We find: $dB^*/df = 2.37$ mT GHz⁻¹, where $B^* = \sum_{j=1}^5 \{ [4/(4j+1)] B_{4/(4j+1)} \} / 5$.

highest f range (Fig. 2b), except that a second zero-resistance state becomes evident around $B = 4/9 B_f$. Here, the sample quality seems to set the lowest magnetic field at which these oscillations can be observed.

Power-dependence data (Fig. 3a) indicate that as the radiation intensity is increased (that is, as the attenuation factor in decibels

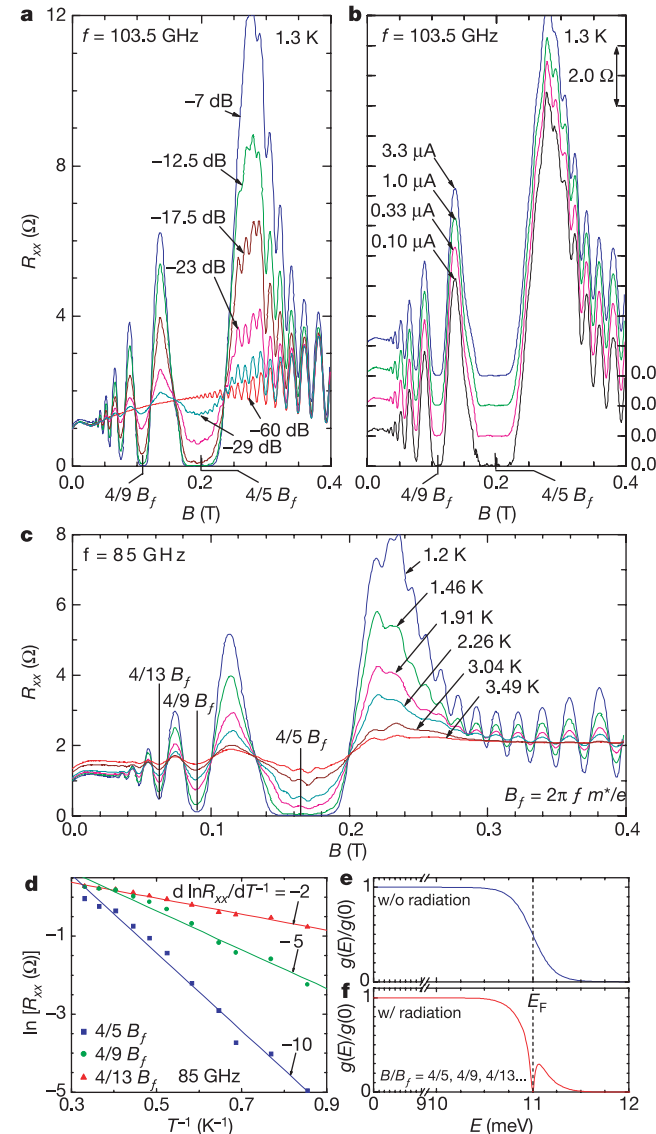


Figure 3 The dependence of the magnetoresistance upon the radiation power, current and the temperature. **a**, The amplitude of the radiation-induced resistance oscillations increases with increasing power, that is, attenuation $\rightarrow 0$ dB, as the resistance at the minima fall below R_{xx} without radiation, leading to zero-resistance states at $4/5 B_f$ and $4/9 B_f$. The radiation power level is estimated to be ≤ 1 mW. **b**, The lineshape of the radiation-induced resistance oscillations is insensitive to the magnitude of the current. **c**, The T -dependence of the magnetoresistance at 85 GHz under constant-power excitation. The radiation-induced resistance minima become deeper at lower temperatures. Note that additional, reproducible, weak resistance oscillations occur about the $4/5 B_f$ minimum at higher temperatures. **d**, The natural logarithm of the resistance versus the inverse temperature, at $B = 4/5 B_f$, $4/9 B_f$ and $4/13 B_f$, for the 85 -GHz data of **c** suggests activated transport and gap formation. Here, $\kappa/2k_B \approx 10$ K for the $4/5 B_f$ state at 85 GHz. **e**, A sketch of the normalized density of states versus the energy in the absence of radiation at low B , when Landau level quantization is imperceptible. **f**, A sketch of the effective density of states versus E about $B = [4/(4j+1)]B_f$, with $j = 1, 2, \dots$ when radiation induces a zero-resistance state. The formation of an energy gap around E_F is suggested by observed activated transport (see **d**).

(dB) $\rightarrow 0$), the resistance at $B = [4/(4j+1)]B_f$ tends to decrease and, about $4/5 B_f$ and $4/9 B_f$, $R_{xx} \rightarrow 0$. The current dependence data shown in Fig. 3b demonstrate insensitivity to the current and the Hall electric field. The temperature (T) variation of R_{xx} at 85 GHz (Fig. 3c) displays both the strong T -dependence of R_{xx} and the low- T requirement for the zero-resistance state. The T -variation of R_{xx} at the deepest minima suggests activated transport; that is, $R_{xx} \propto \exp(-\kappa/2k_B T)$, where k_B is Boltzmann's constant and κ is the

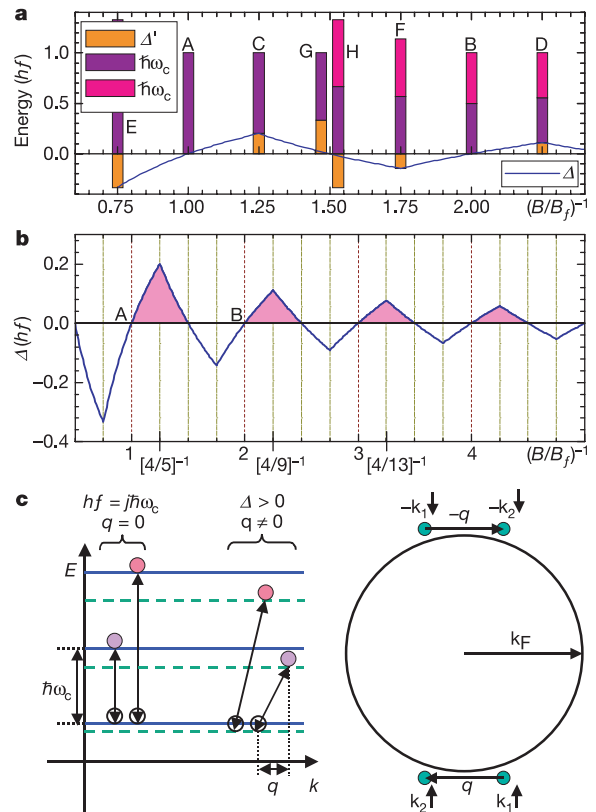


Figure 4 Energy commensurability, inter-Landau-level electron-hole excitations, and the pairing conjecture. **a**, Energy commensurability between the photon energy, hf , and the Landau level spacing, $\hbar\omega_c$, as a function of the normalized inverse field, $(B/B_f)^{-1}$. At $(B/B_f)^{-1}$ marked by A and B, the photon energy hf spans $\hbar\omega_c$ and $2\hbar\omega_c$, respectively, and the energy difference Δ' vanishes, that is, $\Delta' = hf - j\hbar\omega_c = 0$. Small deviations from $(B/B_f)^{-1} = 1$ and $(B/B_f)^{-1} = 2$ lead to deviations from $\Delta' = 0$, as $\hbar\omega_c$ changes with the magnetic field. Thus, at C and D, Δ' becomes positive, and at E and F, Δ' takes on negative values. At $(B/B_f)^{-1} = 3/2$, both $(B/B_f)^{-1} = 1$ and $(B/B_f)^{-1} = 2$ are equidistant in the B^{-1} plot; following the branches from $(B/B_f)^{-1} = 1$ and $(B/B_f)^{-1} = 2$ to $(B/B_f)^{-1} = 3/2$ would yield a double valued Δ' , which is shown as G and H. The B^{-1} periodicity observed in the physical effect, and the symmetry at $(B/B_f)^{-1} = 3/2$, suggest therefore that the physical energy surplus/deficit Δ ought to vanish at $(B/B_f)^{-1} = 3/2$. The full construction of Δ from Δ' follows this notion. **b**, The periodic variation of Δ , the energy surplus/deficit relative to direct inter-Landau-level excitations, is plotted versus $(B/B_f)^{-1}$. The coloured regions represent $(B/B_f)^{-1}$ intervals where Δ is positive. Experimental results (Fig. 1c) indicate that radiation-induced zero-resistance states occur about such $(B/B_f)^{-1}$ intervals. **c**, Left: at $hf = j\hbar\omega_c$, electromagnetic-radiation-induced e-h excitations (excitons) across an integer number of Landau levels do not carry any wavevector, that is, $q = 0$. On the other hand, when $\Delta > 0$, excitons can realize a surplus energy, which can help to access the finite wavevector portion of the excitonic dispersion relation. Right: the Fermi surface in the investigated regime. Under electromagnetic-wave excitation, these specimens can include a steady-state population of e-h excitations (excitons) and a high-mobility two-dimensional 'metal' in a coplanar geometry. q -carrying excitons are proposed to bring about an electron pairing instability by exciton exchange²³⁻²⁷, resulting in a spectral gap. The periodic reappearance of q -carrying excitons leads to the periodic reappearance of such a spectral gap and vanishing resistance.

activation energy (refs 2, 6; Fig. 3d).

Relevant scales include the cyclotron energy, $\hbar\omega_c$, the photon energy, $\hbar f$, and the Landau level broadening $\pi k_B T_D$ (here T_D is the Dingle temperature). At the $B = 0.198$ T minimum (Fig. 1a), $\hbar\omega_c = 0.342$ meV is $4/5$ of $\hbar f = 0.428$ meV at 103.5 GHz, and the Hall effect implies a filling factor of ~ 63 . The transport mean free path is $138 \mu\text{m}$ for $\mu = 1.5 \times 10^7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$. The transport lifetime, $\tau_t = m^* \mu / e = 5.8 \times 10^{-10} \text{ s}$, and the single particle lifetime $\tau_s = \hbar / 2\pi k_B T_D = 1.1 \times 10^{-11} \text{ s}$, suggest a large ratio, $\tau_t / \tau_s \approx 53$, indicative of mostly small angle scattering^{4,10}. The predominance of small angle scattering, and a 7–10 times higher sample mobility, differentiates our specimens from the samples examined in ref. 7. We attribute the occurrence of zero-resistance states in our study, and the lack of them in ref. 7, to these differences. Notably, level broadening determined from T_D suggests that the density of states at low B looks approximately like that at $B = 0$ (Fig. 3e).

In translationally invariant systems, cyclotron resonance is independent of electron–electron interactions and consists of a resonance at $\hbar f = \hbar\omega_c$ only¹³. Our samples include impurity scattering, surface roughness, and a Hall electric field, which could make possible transitions at $\hbar f = j\hbar\omega_c$ (refs 14, 15). In a bounded specimen, the collective plasma mode can also hybridize with cyclotron resonance, yielding magnetoplasmons at $B > 0$ (refs 16, 17). A search for plasma frequency, ω_p (ref. 16), sensitivity produced a null result^{5,19}. In addition, R_{xx} under radiation does not directly manifest the bare cyclotron resonance ($\hbar\omega_c$), or its harmonics ($j\hbar\omega_c$), as evidenced by the phase shift of the extrema with respect to integral B_f^{-1} in a B^{-1} plot. Thus, there is no obvious violation of Kohn's theorem¹³.

The sample mobility ($\mu \geq 10^7 \text{ cm}^2 \text{ V}^{-1} \text{ s}^{-1}$) implies that $\mu B > 1$ for $B \geq 1$ mT. Thus, where R_{xx} vanishes, the Hall angle is $\sim 90^\circ$, implying electron transport along equipotentials²⁴. In this limit, back scattering at the Fermi surface yields a finite R_{xx} . The radiation-induced vanishing-resistance thus indicates that scattering is prohibited, while the activated resistance indicates that gap formation (Fig. 3f) suppresses scattering.

A full theory of this phenomenon, which considers the interaction of electrons with the electromagnetic field, awaits development. Meanwhile, we suggest an approach that relates a spectral gap to the presence of radiation-induced excitons, via an excitonic mechanism for electron pairing: in this 2DES, excitonic electron–hole (e–h) excitations between Landau levels near the Fermi level can be produced by photons, upon conserving energy and wavevector, when $\hbar f$ approaches $j\hbar\omega_c$ (refs 20–22). Further, a steady-state population of excitons can occur under constant radiation. Thus, these specimens could include electrons and excitons in a coplanar geometry, resembling proposals where a weak attractive interaction between electrons can occur by exciton exchange^{23–27}. Unlike some considered systems^{23–26}, this 2DES includes low-energy excitons which implies gap formation at low ($\sim \hbar\omega_c / k_B$) temperatures²⁴. Yet, the periodicity and phase of the zero-resistance states, the role of B_f and the power dependence can be qualitatively understood within such an approach, after accounting for Landau quantization.

According to theory, excitons in the 2DES follow the energy dispersion relation $E_\mu^j = j\hbar\omega_c + \Delta E_\mu^j(q)$ and include a wavevector q -dependent energy shift of order $e^2 / \epsilon l_0$, where l_0 is the magnetic length^{21,22}. Thus, a necessary component for exciton-mediated electron pairing, finite- q excitons²⁵, might be accessible to the system when electron–hole excitations obtain a surplus energy ΔE_μ^j . We examine the interplay between the fixed $\hbar f$ and the B -tunable $\hbar\omega_c$ (Fig. 4a) in order to identify the regime where a surplus energy might occur. In this figure, at A and B, an integer number of Landau levels are commensurate with $\hbar f$, and the energy difference vanishes, $\Delta' = \hbar f - j\hbar\omega_c = 0$, indicating the absence of an energy surplus or deficit. Also, small excursions from $(B/B_f)^{-1} = 1$ (or 2) lead to deviations from $\Delta' = 0$. Following the branches from $(B/B_f)^{-1} = 1$ and $(B/B_f)^{-1} = 2$ yields a double valued Δ' at

$(B/B_f)^{-1} = 3/2$. The B^{-1} periodicity observed in the resistance, and the symmetry at $(B/B_f)^{-1} = 3/2$, suggest that the physical energy surplus/deficit, Δ , ought to vanish at $(B/B_f)^{-1} = 3/2$ in this situation. The construction of Δ from Δ' (Fig. 4b) follows this notion. Then, Fig. 4b indicates that Δ takes on maximal positive values at $(B/B_f)^{-1} = [4/(4j+1)]^{-1}$, favouring finite- q excitons over the interval $[4/4j]^{-1} < (B/B_f)^{-1} < [4/(4j+2)]^{-1}$. The data (Fig. 1c) indicate that radiation reduces R_{xx} over corresponding intervals.

When $\Delta > 0$, electron pairing by exciton exchange can occur as in Fig. 4c (ref. 25). Here, the scattering of an electron from $\mathbf{k}_1$ to $\mathbf{k}_2 = \mathbf{k}_1 + \mathbf{q}$ provides wavevector for the exciton, and facilitates access to the finite- q portion of the dispersion relation. Such correlated scattering could produce a pairing instability near the Fermi surface^{23–27}, resulting in a spectral gap. As radiation facilitates q -carrying excitons, a spectral gap can occur only under illumination. The periodic (in B^{-1}) reappearance of q -carrying excitons leads to the periodic reoccurrence of electron pairing, and vanishing resistance. In contrast, when $\Delta < 0$, a local energy deficit for exciton creation increases the electron inelastic-scattering cross-section, producing radiation-enhanced resistance over corresponding intervals. Thus, this interpretation suggests that excitons induce a pairing instability^{23–26} in the 2DES system around $B = [4/(4j+1)]B_f$, which leads to a spectral gap, activated transport⁶ and zero-resistance-states³, while the lack of phase coherence prevents the realization of typical superconductivity^{1,28}. These results complement studies of exciton condensation in the quantum Hall regime^{29,30}, and studies of exciton ordering and transport in coupled quantum wells^{31,32}. □

Received 10 June; accepted 4 November 2002; doi:10.1038/nature01277.

1. Tinkham, M. *Introduction to Superconductivity*, 2nd edn (McGraw-Hill, New York, 1996).
2. Prange, R. E. & Girvin, S. M. (eds) *The Quantum Hall Effect*, 2nd edn (Springer, New York, 1990).
3. Tsui, D. C., Stormer, H. L. & Gossard, A. C. Zero-resistance state of two dimensional electrons in a quantizing magnetic field. *Phys. Rev. B* **25**, 1405–1407 (1982).
4. Ando, T., Fowler, A. B. & Stern, F. Electronic properties of two-dimensional systems. *Rev. Mod. Phys.* **54**, 437–672 (1982).
5. Mani, R. G., Smet, J. H., von Klitzing, K., Narayanamurti, V. & Umansky, V. Single particle and collective response in the magnetophotoreistance of a high mobility 2DES under microwave excitation. *Bull. Am. Phys. Soc.* **46**, 972 (2001).
6. von Klitzing, K., et al. *Proc. 17th Int. Conf. on the Physics Of Semiconductors* (eds Chadi, D. J. & Harrison, W. A.) 271–274 (Springer, New York, 1985).
7. Zudov, M. A., Du, R. R., Simmons, J. A. & Reno, J. L. Shubnikov-de Haas-like oscillations in millimeterwave photoconductivity in a high mobility two-dimensional electron gas. *Phys. Rev. B* **64**, 201311 (2001).
8. Nicholas, R. J., Brummell, M. A. & Portal, J. C. *Two-Dimensional Systems, Heterostructures and Superlattices* Springer Series in Solid State Sciences (eds Bauer, G., Kuchar, F. & Heinrich, H.) Vol. 53 69–78 (Springer, Berlin, 1984).
9. Sze, S. M. *Physics of Semiconductor Devices*, 2nd edn 850 (Wiley, New York, 1981).
10. Mani, R. G. & Anderson, J. R. Study of the single particle and transport lifetimes in GaAs/Al_xGa_{1-x}As. *Phys. Rev. B* **37**, 4299–4302 (1988).
11. Englert, Th., Mann, J. C., Uihlein, Ch., Tsui, D. C. & Gossard, A. C. Observation of oscillatory linewidth in the cyclotron resonance of GaAs/Al_xGa_{1-x}As heterostructures. *Solid State Commun.* **46**, 545–548 (1983).
12. Schlesinger, Z., Allen, S. J., Huang, J. C. M., Platzmann, P. M. & Tzoar, N. Cyclotron resonance in two dimensions. *Phys. Rev. B* **30**, 435–437 (1984).
13. Kohn, W. Cyclotron resonance and de Haas-van Alphen oscillations of an interacting electron gas. *Phys. Rev.* **123**, 1242–1242 (1961).
14. Ando, T. Mass enhancement and subharmonic structure of cyclotron resonance in an interacting two-dimensional electron gas. *Phys. Rev. Lett.* **36**, 1383–1385 (1976).
15. Klahn, S., Horst, M. & Merkt, U. *Proc. 18th Int. Conf. on the Physics of Semiconductors* (ed. Engstrom, O.) Vol. 2 1161–1163 (World Scientific, Singapore, 1987).
16. Heitmann, D. Two-dimensional plasmons in homogeneous and laterally microstructured space charge layers. *Surf. Sci.* **170**, 332–345 (1986).
17. Allen, S. J., Jr, Tsui, D. C. & Logan, R. A. Observation of two-dimensional plasmon in silicon inversion layers. *Phys. Rev. Lett.* **38**, 980–983 (1977).
18. Theis, T. N., Kotthaus, J. P. & Stiles, P. J. Wavevector dependence of the two-dimensional plasmon dispersion relationship in the (100) Si inversion layer. *Solid State Commun.* **26**, 603–606 (1977).
19. Vasiladoulou, E. et al. Collective response in the microwave photoconductivity of Hall bar structures. *Phys. Rev. B* **48**, 17145–17148 (1993).
20. Lerner, I. V. & Lozovik, Yu. E. Two-dimensional electron-hole system in a strong magnetic field as an almost ideal exciton gas. *Sov. Phys. JETP* **53**, 763–770 (1981).
21. Kallin, C. & Halperin, B. I. Excitations from a filled Landau level in the two-dimensional electron gas. *Phys. Rev. B* **30**, 5655–5668 (1984).
22. Bychkov, Yu. A., Maniv, T. & Vagner, I. D. Nuclear spin diffusion via spin-excitons in the quantum Hall regime. *Solid State Commun.* **94**, 61–65 (1995).
23. Little, W. A. Possibility of synthesizing an organic superconductor. *Phys. Rev.* **134**, A1416–A1424 (1965).
24. Ginzburg, V. L. The problem of high temperature superconductivity. II. *Sov. Phys. Uspekhi.* **13**, 335–352 (1970).

25. Allender, D., Bray, J. & Bardeen, J. Model for an exciton mechanism of superconductivity. *Phys. Rev. B* **7**, 1020–1029 (1973).
26. Davis, D., Gutfreund, H. & Little, W. A. Proposed model of a high temperature excitonic superconductor. *Phys. Rev. B* **13**, 4766–4779 (1976).
27. Hirsch, J. E. & Scalapino, D. J. Enhanced superconductivity in quasi two-dimensional systems. *Phys. Rev. Lett.* **56**, 2732–2735 (1986).
28. Bezryadin, A., Lau, C. N. & Tinkham, M. Quantum suppression of superconductivity in ultrathin nanowires. *Nature* **404**, 971–974 (2000).
29. Paquet, D., Rice, T. M. & Ueda, K. Two-dimensional electron-hole fluid in a strong perpendicular magnetic field: exciton Bose condensate or maximum density two-dimensional droplet. *Phys. Rev. B* **32**, 5208–5221 (1985).
30. Kellogg, M., Spielman, I. B., Eisenstein, J. P., Pfeiffer, L. N. & West, K. W. Observation of quantized Hall drag in a strongly correlated bilayer electron system. *Phys. Rev. Lett.* **88**, 126804 (2002).
31. Butov, L. V., Gossard, A. C. & Chemla, D. S. Macroscopically ordered state in an exciton system. *Nature* **418**, 751–754 (2002).
32. Snoke, S., Denev, S., Liu, Y., Pfeiffer, L. & West, K. Long-range transport in excitonic dark states in coupled quantum wells. *Nature* **418**, 754–757 (2002).

Acknowledgements We acknowledge discussions with E. Demler, H. Fertig, R. Gerhardt, W. Hanke, C. Kallin, M. Kruger, L. Manchanda, S. Mikhailov, A. Stern and M. Tinkham. This work has been supported by the ARO, BMBF, CSR at SRC, DFG and GIF.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to R.G.M. (e-mail: mani@deas.harvard.edu).

Wavelength-scalable hollow optical fibres with large photonic bandgaps for CO₂ laser transmission

Burak Temelkuran*, Shandon D. Hart*, Gilles Benoit, John D. Joannopoulos & Yoel Fink

Research Laboratory of Electronics and Department of Materials Science and Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA

* These authors contributed equally to this work

Conventional solid-core optical fibres require highly transparent materials. Such materials have been difficult to identify owing to the fundamental limitations associated with the propagation of light through solids, such as absorption, scattering and nonlinear effects. Hollow optical fibres offer the potential to minimize the dependence of light transmission on fibre material transparency^{1–3}. Here we report on the design and drawing of a hollow optical fibre lined with an interior omnidirectional dielectric mirror⁴. Confinement of light in the hollow core is provided by the large photonic bandgaps^{5–7} established by the multiple alternating submicrometre-thick layers of a high-refractive-index glass and a low-refractive-index polymer. The fundamental and high-order transmission windows are determined by the layer dimensions and can be scaled from 0.75 to 10.6 μm in wavelength. Tens of metres of hollow photonic bandgap fibres for transmission of carbon dioxide laser light at 10.6 μm wavelength were drawn. The transmission losses are found to be less than 1.0 dB m⁻¹, orders of magnitude lower than those of the intrinsic fibre material, thus demonstrating that low attenuation can be achieved through structural design rather than high-transparency material selection.

Silica optical fibres^{8,9} have been extremely successful in telecommunications applications, and other types of solid-core fibres have been explored at wavelengths where silica is not transparent^{10,11}. However, all fibres that rely on light propagation principally through a solid material have certain fundamental limitations stemming from nonlinear effects, light absorption by electrons or

phonons, material dispersion and Rayleigh scattering that limit maximum optical transmission power and increase attenuation losses^{12,13}. These limitations have in turn motivated the study of a fundamentally different light-guiding methodology: the use of hollow waveguides having highly reflecting walls. Light propagation through air in a hollow fibre eliminates or greatly reduces the problems of nonlinearities, thermal lensing and end-reflections, facilitating high-power laser guidance and other applications that may be impossible using conventional fibres^{12,14,15}. Hollow metallic or metallo-dielectric waveguides^{12,16–19} have been studied fairly extensively and found useful practical applications, but their performance has been bounded by the notable losses occurring in metallic reflections at visible and infrared wavelengths, as well as by the limited length and mechanical flexibility of the fabricated waveguides. Hollow all-dielectric fibres relying on specular or attenuated total reflection have also been explored, but high transmission losses have prevented their broad application^{12,20,21}. More recently, all-dielectric fibres consisting of a periodic array of air holes in silica have been used to guide light through air using narrow photonic bandgaps^{1–3}. Solid-core, index guiding versions of these silica photonic crystal fibres have also been explored for applications such as very-large-core single-mode fibres, nonlinear enhancement and broadband supercontinuum generation, polarization maintenance and dispersion management³. However, the air-guiding capabilities of such waveguides have remained limited thus far by the difficulties in fabricating long, uniform fibres which must have a high volume fraction of air and many air-hole periods, as well as by the large electromagnetic penetration depths associated

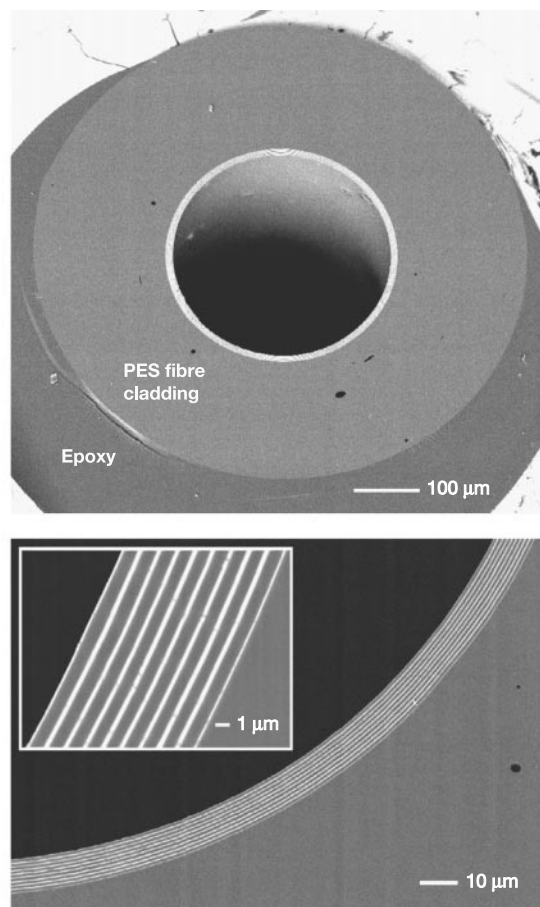


Figure 1 Cross-sectional SEM micrographs at various magnifications of hollow cylindrical multilayer fibre mounted in epoxy. The hollow core appears black, the PES layers and cladding grey, and the As₂Se₃ layers bright white. This fibre has a fundamental photonic bandgap at a wavelength of $\sim 3.55 \mu\text{m}$.

with the small photonic bandgaps achievable in these air–silica structures^{2,22}.

In the fibre we report here, the hollow core is surrounded by a solid multilayer structure of high refractive-index contrast, leading to large photonic bandgaps and omnidirectional reflectivity. The pertinent theoretical background²³ and recent analyses^{24–27} indicate that such fibres may be able to achieve ultralow losses and other unique transmission properties (M. Ibanescu *et al.*, manuscript in preparation). The large photonic bandgaps result in very short electromagnetic penetration depths within the layer structure, significantly reducing radiation and absorption losses while increasing robustness.

To achieve high index contrast in the layered portion of the fibre, we combined a chalcogenide glass with a refractive index of ~ 2.8 , arsenic triselenide (As_2Se_3), with a high glass-transition temperature thermoplastic polymer having a refractive index of ~ 1.55 , poly(ether sulphone) (PES)³⁰. We recently demonstrated that these materials could be thermally co-drawn into precisely layered structures without cracking or delamination, even under large temperature excursions²⁸. The same polymer was used as a cladding material, resulting in fibres composed of $\sim 98\%$ polymer by volume (not including the hollow core); the fibres thus combine high optical performance with polymeric processability and mechanical flexibility. We fabricated a variety of fibres by depositing an As_2Se_3 layer (5–10 μm thick) by thermal evaporation onto a 25–50- μm -thick PES film, and the subsequent ‘rolling’ of that coated film into a hollow multilayer tube called a fibre pre-form. This hollow macroscopic pre-form was consolidated by heating under vacuum, and clad with a thick outer layer of PES; the layered pre-form was then placed in an optical fibre draw tower, and drawn down into tens or hundreds of metres of fibre having well-controlled submicrometre layer thicknesses. The nominal positions of the photonic bandgaps were determined by laser monitoring of the fibre outer diameter during the draw process. Typical standard

deviations in the fibre outer diameter were $\sim 1\%$ of that diameter. The resulting fibres were designed to have large hollow cores, useful in high-energy transmission.

Scanning electron microscope (SEM) analysis (Fig. 1) reveals that the drawn fibres maintain proportionate layer thickness ratios, and that the PES and As_2Se_3 films adhere well during rigorous thermal cycling and elongation. Within the multilayer structure shown in Fig. 1, the PES layers (grey) have a thickness of 900 nm, and the As_2Se_3 layers (bright) are 270 nm thick (except for the first and last As_2Se_3 layers, which are 135 nm). Broadband fibre transmission spectra were measured with a Fourier-transform infrared (FTIR) spectrometer (Nicolet Magna 860), using a parabolic mirror to couple light into the fibre and an external detector. The results of these measurements are shown in Fig. 2, bottom panel, for fibres having two different layer structures. For each spectrum, light is guided at the fundamental and high-order photonic bandgaps. The top panel of Fig. 2 shows the corresponding photonic band diagram^{4,7} for an infinite periodic multilayer structure, calculated using the experimental parameters of our fibre (layer thicknesses and indices). Good agreement is found between the positions of the measured transmission peaks and the calculated bandgaps, corroborated by the SEM-measured layer thicknesses, verifying that transmission is dominated by the photonic bandgap mechanism.

In order to demonstrate ‘wavelength scalability’ (that is, the control of transmission through the fibre’s structural parameters), another fibre was produced having the same cross-section but with thinner layers. We compared the transmission spectrum of the original 3.55- μm -bandgap fibres with that of the fibre having scaled-down layer thicknesses. Figure 2 shows the shifting of the transmission bands, corresponding to fundamental and high-order photonic bandgaps, from one fibre to the next. The two fibres analysed in Fig. 2 were fabricated from the same fibre pre-form using different draw-down ratios (fibres with a fundamental bandgap centred near 3.55 μm have an outer diameter of 670 μm ; those with a gap at 3.1 μm have an outer diameter of 600 μm). The high-order bandgaps are periodically spaced in frequency, as expected for such a photonic crystal structure. The fibre having a fundamental bandgap at 3.1 μm has a fourth-order bandgap at 775 nm (measured using a StellarNet EPP2000 spectrometer); this output from the hollow fibre core is imaged in Fig. 3 inset. Broadband light from a quartz–tungsten–halogen lamp was coupled into the 275- μm -diameter hollow core of the fibre shown in Fig. 3. Owing to the proximity of this wavelength to the electronic band edge of the As_2Se_3 glass, extra precautions were needed to reduce transmission due to glancing angle specular reflection. By placing multiple bends in the fibre, light transmitted owing to this effect was dramatically

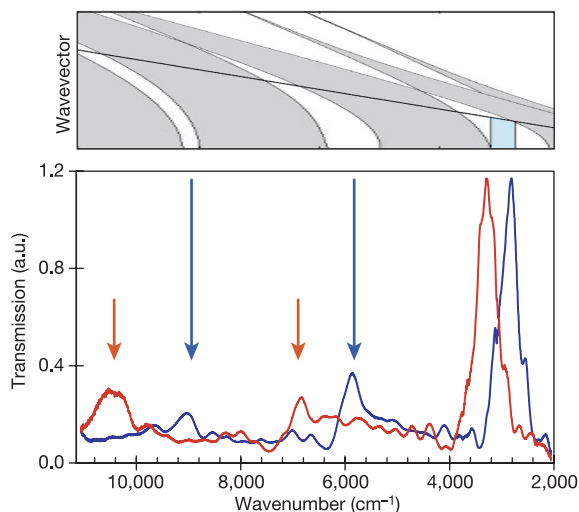


Figure 2 Photonic band structure due to the dielectric mirror, and the resulting transmission spectra for the hollow fibre. Upper panel, calculated photonic band structure associated with the dielectric mirror lining of the hollow fibre. Modes propagating through air and reflected by the fibre walls lie in the bandgaps (white) and within the light cone defined by the glancing-angle condition (black line). The grey regions represent modes radiating through the mirror. The fundamental bandgap has the widest range-to-midrange ratio, with a region of omnidirectional reflectivity highlighted in blue. Lower panel, comparison of transmission spectra for two different hollow fibres of ~ 30 cm length having similar structure but scaled photonic crystal (multilayer) period dimensions. The spectrum in blue is from a fibre with a fundamental photonic bandgap at 3.55 μm ; the red spectrum is from a fibre where the corresponding bandgap is at 3.1 μm . High-order bandgaps (indicated by arrows) are periodically spaced in frequency.

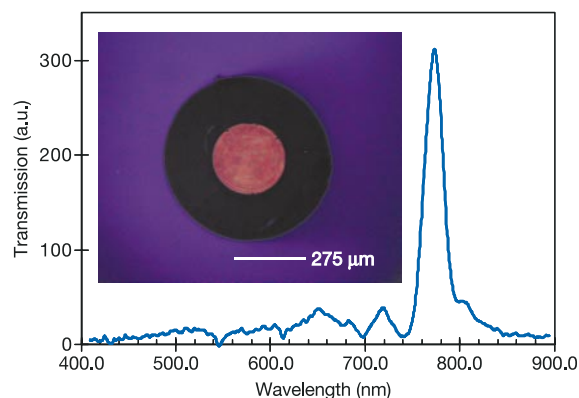


Figure 3 Visible to near-infrared transmission spectrum and charge-coupled device (CCD) image (inset) of light emerging from core of hollow fibre that has a fundamental bandgap at 3.1 μm . This fibre was ~ 25 cm long and bent multiple times. Imaged light is confined to the hollow core by the fourth-order photonic bandgap.

decreased. In short fibres (~ 30 cm), shallow bending was not found to greatly reduce transmission in the fundamental photonic bandgap, though bending losses in the high-order bandgaps are much greater. This should be expected, because the fundamental gap exhibits omnidirectional reflectivity while the high-order gaps do not. In order to qualitatively test the limits of this behaviour, a ~ 30 -cm fibre having a bandgap at $3.55\text{ }\mu\text{m}$ was looped into a small 'knot' consisting of multiple bends (Fig. 4). Approximately 40% of the light in the fundamental bandgap was transmitted through this fibre when compared to the straight fibre, while the light guided in the high-order gaps was attenuated much more strongly. Hollow fibres with bandgaps near $3\text{ }\mu\text{m}$ are of interest for the transmission of high-energy Er:YAG laser radiation.

The wavelength scalability of our fibres was further demonstrated by the fabrication of hollow fibres designed for the transmission of $10.6\text{-}\mu\text{m}$ electromagnetic radiation. This not only shows that these structures can be made to guide light at extremely disparate wavelengths, but that specific bandgap wavelengths can be accurately targeted during fabrication and fibre drawing. Powerful and efficient CO_2 lasers are available that emit at $10.6\text{ }\mu\text{m}$, and are used in such applications as laser surgery and materials processing, but waveguides operating at this wavelength have remained limited in length or loss levels^{12,18}. Using the fabrication techniques outlined above, we produced fibres having hollow core diameters of $700\text{--}750\text{ }\mu\text{m}$ and outer diameters of $1,300\text{--}1,400\text{ }\mu\text{m}$, with a fundamental photonic bandgap spanning the $10\text{--}11\text{ }\mu\text{m}$ wavelength regime, centred near $10.6\text{ }\mu\text{m}$. Figure 5 depicts a typical FTIR transmission spectrum for these fibres, measured using ~ 30 -cm-long straight fibres.

In order to quantify the transmission losses in these $10.6\text{-}\mu\text{m}$ -bandgap hollow fibres, fibre cutback measurements were performed. This involved the comparison of transmitted intensity through $\sim 4\text{ m}$ of straight fibre with the intensity of transmission through the same section of fibre cut to shorter lengths (Fig. 5 inset). This test was performed on multiple sections of fibre, and the results found to be nearly identical for the different sections tested. The measurements were performed using a 25 W CO_2 laser (GEM-25, Coherent-DEOS) and high-power detectors (Newport 818T-10). The fibre was held straight; it was fixed at both ends, as well as at points near the middle to prevent variations in the input coupling and propagation conditions during fibre cutting. The laser beam was sent through focusing lenses as well as $500\text{-}\mu\text{m}$ -diameter pin-hole apertures, and the input end face of the fibre was coated with a metal film to prevent accidental laser damage from misalignment (see Supplementary Information for further measurement details).

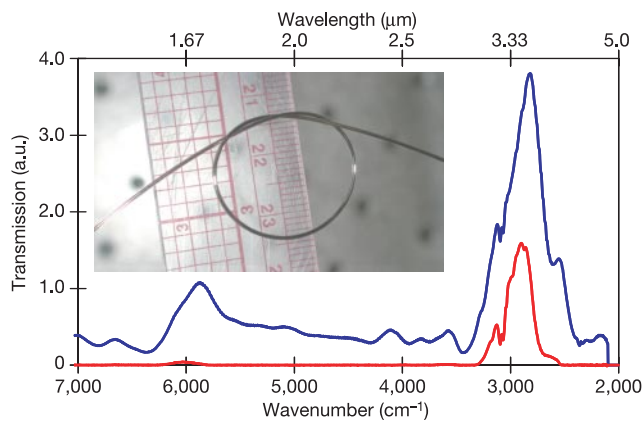


Figure 4 Transmission spectra for straight (blue) and 'knotted' (red) hollow fibre having a fundamental bandgap at $3.55\text{ }\mu\text{m}$. The fibre knot consisted of multiple bends and the knot had a radius of $\sim 1\text{ cm}$ (see inset). Approximately 40% of the light in the fundamental bandgap was transmitted through this highly perturbed fibre, whereas the high-order bandgap at $1.7\text{ }\mu\text{m}$ has much greater bending losses.

The transmission losses in the fundamental bandgap at $10.6\text{ }\mu\text{m}$ were measured to be 0.95 dB m^{-1} , as shown in the inset of Fig. 5, with an estimated measurement uncertainty of 0.15 dB m^{-1} . These loss measurements are comparable to some of the best reported loss values for other types of waveguides operating at $10.6\text{ }\mu\text{m}$ (refs 12, 29). A bending analysis for fibres with a bandgap centred at $10.6\text{ }\mu\text{m}$ revealed bending losses below 1.5 dB for 90° bends with bending radii from 4 to 10 cm (see Supplementary Information for further discussion of bending losses as well as modal considerations). We expect that these loss levels could be lowered even further by increasing the number of layers, through optimization of the layer thickness ratios and by creating a cylindrically symmetric multilayer fibre with no inner seam (present here because of the 'rolling' fabrication method). In addition, using a polymer with lower intrinsic losses should greatly improve the transmission characteristics.

One reasonable figure of merit for optical transmission losses through hollow all-dielectric photonic bandgap fibres is a comparison of the hollow fibre losses to the intrinsic losses of the materials used to make the fibre. As_2Se_3 has been explored as an infrared-transmitting material, yet the losses at $10.6\text{ }\mu\text{m}$ reported in the literature are $\sim 7\text{ dB m}^{-1}$ for highly purified material, and more typically are greater than 10 dB m^{-1} for commercially available materials such as those used in our fabrication²⁹. Based on FTIR transmission and spectroscopic ellipsometer measurements that we have performed on PES, the optical losses associated with propagation through solid PES should be greater than $100,000\text{ dB m}^{-1}$ at $10.6\text{ }\mu\text{m}$. This demonstrates that guiding light through air in our hollow bandgap fibres leads to waveguide losses that are orders of magnitude lower than the intrinsic fibre material losses, which has been one of the primary goals of research into hollow photonic bandgap fibres. These comparatively low losses are made possible by the very short penetration depths of electromagnetic waves in the high-refractive-index-contrast photonic crystal structure, allowing these materials to be used at wavelengths that may have been thought improbable¹⁸.

Another long-standing motivation of infrared fibre research has been the transmission of high-power laser light^{11,19}. As a qualitative demonstration of the potential of our fibres for such applications, both straight and smoothly bent fibres of lengths 0.3 to 2.5 m were used to transmit sufficient CO_2 laser energy to burn holes in paper and a film of PES (the fibre majority component). This demonstrates the substantial reduction of fibre transmission loss relative to the intrinsic materials loss of the constituents (see Supplementary Information). The maximum laser power density coupled into our fibres in these trials was approximately 300 W cm^{-2} . No damage to

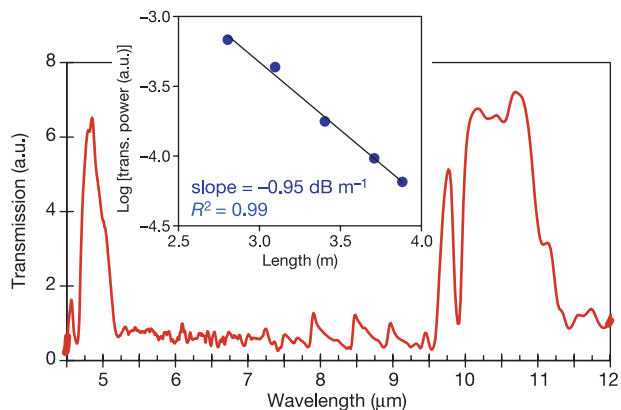


Figure 5 Typical transmission spectrum of hollow fibres designed to transmit CO_2 laser light. The fundamental photonic bandgap is centred near a wavelength of $10.6\text{ }\mu\text{m}$, and the second-order gap is at $\sim 5\text{ }\mu\text{m}$. Inset, $\log(\text{transmitted power})$ versus length of fibre. The slope of this graph is the loss in dB m^{-1} . The measured fibre has a hollow core diameter of $700\text{ }\mu\text{m}$.

the fibres was observed when the laser beam was properly coupled into the hollow fibre core. These results indicate the feasibility of using hollow multilayer photonic bandgap fibres as a low-loss wavelength-scalable transmission medium for high-power laser light. □

Received 3 July; accepted 29 October 2002; doi:10.1038/nature01275.

1. Cregan, R. F. *et al.* Single-mode photonic band gap guidance of light in air. *Science* **285**, 1537–1539 (1999).
2. Allan, D. C. *et al.* *Photonic Crystals and Light Localization in the 21st Century* (ed. Soukoulis, C. M.) 305–320 (Kluwer, Boston, 2001).
3. Eggleton, B. J., Kerbage, C., Westbrook, P. S., Windeler, R. S. & Hale, A. Microstructured optical fiber devices. *Opt. Express* **9**, 698–713 (2001).
4. Fink, Y. *et al.* A dielectric omnidirectional reflector. *Science* **282**, 1679–1682 (1998).
5. Yablonovitch, E. Inhibited spontaneous emission in solid-state physics and electronics. *Phys. Rev. Lett.* **58**, 2059–2062 (1987).
6. John, S. Strong localization of photons in certain disordered dielectric superlattices. *Phys. Rev. Lett.* **58**, 2486–2489 (1987).
7. Joannopoulos, J. D., Meade, R. D. & Winn, J. N. *Photonic Crystals: Molding the Flow of Light* (Princeton Univ. Press, Princeton, New Jersey, 1995).
8. Maurer, R. D. & Schultz, P. C. US Patent 3,659,915 (1972).
9. Keck, D. B., Maurer, R. D. & Schultz, P. C. On the ultimate lower limit of attenuation in glass optical waveguides. *Appl. Phys. Lett.* **22**, 307–309 (1973).
10. Hilton, A. R. Optical properties of chalcogenide glasses. *J. Non-Cryst. Solids* **2**, 28–39 (1970).
11. Harrington, J. A. *Handbook of Optics* (ed. Bass, M.) 14.1–14.13 (McGraw-Hill, New York, 2001).
12. Harrington, J. A. A review of IR transmitting, hollow waveguides. *Fiber Integr. Opt.* **19**, 211–227 (2000).
13. Mitra, P. P. & Stark, J. B. Nonlinear limits to the information capacity of optical fibre communications. *Nature* **411**, 1027–1030 (2001).
14. Renn, M. J. *et al.* Laser-guided atoms in hollow-core optical fibers. *Phys. Rev. Lett.* **75**, 3253–3256 (1995).
15. Rundquist, A. *et al.* Phase-matched generation of coherent soft X-rays. *Science* **280**, 1412–1415 (1998).
16. Marcatilli, E. A. J. & Schmeltzer, R. A. Hollow metallic and dielectric waveguides for long distance optical transmission and lasers. *Bell Syst. Tech. J.* **43**, 1783–1809 (1964).
17. Miyagi, M. & Kawakami, S. Design theory of dielectric-coated circular metallic waveguides for infrared transmission. *J. Lightwave Technol.* **2**, 116–126 (1984).
18. Matsuura, Y., Kasahara, R., Katagiri, T. & Miyagi, M. Hollow infrared fibers fabricated by glass-drawing technique. *Opt. Express* **10**, 488–492 (2002).
19. Hongo, A., Morosawa, K., Matsumoto, K., Shiota, T. & Hashimoto, T. Transmission of kilowatt-class CO₂-laser light through dielectric-coated metallic hollow wave-guides for material processing. *Appl. Opt.* **31**, 5114–5120 (1992).
20. Bornstein, A. & Croitoru, N. Chalcogenide hollow fibers. *J. Non-cryst. Solids* **77–78**, 1277–1280 (1985).
21. Fitt, A. D., Furusawa, K., Monro, T. M. & Please, C. P. Modeling the fabrication of hollow fibers: Capillary drawing. *J. Lightwave Technol.* **19**, 1924–1931 (2001).
22. Broeng, J., Barkou, S. E., Sondergaard, T. & Bjarklev, A. Analysis of air-guiding photonic bandgap fibers. *Opt. Lett.* **25**, 96–98 (2000).
23. Yeh, P., Yariv, A. & Marom, E. Theory of Bragg fiber. *J. Opt. Soc. Am.* **68**, 1196–1201 (1978).
24. Ouyang, G., Xu, Y. & Yariv, A. Comparative study of air-core and coaxial Bragg fibers: single-mode transmission and dispersion characteristics. *Opt. Express* **9**, 733–747 (2001).
25. Ibanescu, M., Fink, Y., Fan, S., Thomas, E. L. & Joannopoulos, J. D. An all-dielectric coaxial waveguide. *Science* **289**, 415–419 (2000).
26. Johnson, S. G. *et al.* Low-loss asymptotically single-mode propagation in large-core OmniGuide fibers. *Opt. Express* **9**, 748–779 (2001).
27. Fink, Y. *et al.* Guiding optical light in air using an all-dielectric structure. *J. Lightwave Technol.* **17**, 2039–2041 (1999).
28. Hart, S. D. *et al.* External reflection from omnidirectional dielectric mirror fibers. *Science* **296**, 510–513 (2002).
29. Sanghera, J. S. & Aggarwal, I. D. Active and passive chalcogenide glass optical fibers for IR applications: a review. *J. Non-cryst. Solids* **257**, 6–16 (1999).
30. Bormashenko, E., Pogreb, R., Pogreb, Z. & Sutoviski, S. Development of new near-infrared filters based on the “sandwich” polymer-chalcogenide glass-polymer composites. *Opt. Eng.* **40**, 661–662 (2001).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank P. H. Pridaux for teaching us the ways and means of optical fibre drawing; G. R. Maskaly, H. Burch, K. R. Maskaly, E. P. Chan, O. Shapira, M. Bayindir, C. H. Sarantos and C. Guequeta for their contributions; L. H. Galindo, T. McClure and M. Frongillo for experimental aid; W. A. King, J. A. Harrington, A. R. Hilton, E. L. Thomas, U. Kolodny and R. Stata for discussions and support; L. Laughman, L. Newman, A. DeMaria and the team at Coherent-DEOS for assistance; and W. H. Smith, M. Young and the MIT-RLE for administrative support. This work was supported in part by DARPA-QUIST/ARO, the NSF, the US DOE, and an NSF graduate research fellowship (S.D.H.). This work was also supported by the Materials Research Science and Engineering Center (MRSEC) programme of the NSF, and made use of MRSEC shared facilities.

Competing interests statement The authors decline to provide information about competing financial interests.

Correspondence and requests for materials should be addressed to Y.F.
(e-mail: yoel@mit.edu).

Bond-controlled configurational entropy reduction in chemical vitrification

Silvia Corezzi*, Daniele Fioretto* & Pierangelo Rolla†

* Istituto Nazionale per la Fisica della Materia and Dipartimento di Fisica, Università di Perugia, Via A. Pascoli, I-06123, Perugia, Italy

† Istituto Nazionale per la Fisica della Materia and Dipartimento di Fisica, Università di Pisa, Via F. Buonarroti 2, I-56127, Pisa, Italy

Glass formation is usually viewed in terms of physical vitrification: a liquid in a metastable state¹ is cooled or compressed so as to avoid crystallization. However, glasses may also be formed by chemical vitrification, a process involving progressive polymerization of the constituent molecules via the formation of irreversible chemical bonds. The formation of most of the materials used in engineering plastics and the hardening of natural and synthetic resins are based on chemical vitrification. Despite the differences in the molecular processes involved in chemical and physical vitrification, surprising similarities^{2–9} are observed in the slowing down of the dynamics and in the thermodynamical properties of the resulting glasses. Explaining such similarities would improve general understanding of the glass transition and may disclose its universal nature. Here we report dielectric and photon-correlation measurements that reveal the origin of the similarity in the dynamical behaviour of physical and chemical glass formers. We find that the evolution of their configurational restrictions proceeds in a similar manner. In particular, we make a connection between the reduction in configurational entropy and the number of chemical bonds, a quantity that can be controlled in experiments.

The polymerization of liquid monomers is an irreversible process that modifies molecular motions in reacting systems in a manner that closely resembles the reversible effect of reducing temperature or increasing pressure in chemically stable glass-forming liquids, and ultimately leads to a ‘jammed’ structure. However, glass science has mainly focused on glasses obtained following changes of temperature and pressure, called here the ‘physical’ glasses. No comparable effort has been made to clarify how microscopic processes of a chemical nature can result in a similar state of matter. In principle, there exists the possibility that physical and chemical vitrification have nothing in common apart from the final state. We show here that this is not the case.

In physical glasses, the dynamics of liquids approaching the glass transition strongly reflect the underlying thermodynamics of the liquid state. One way of rationalizing this qualitative notion is in terms of the Adam–Gibbs theory¹⁰, establishing a link between the structural relaxation time τ (a dynamic quantity) and the configurational entropy S_c (a thermodynamic quantity) of a liquid, in the form

$$\tau = \tau_0 \exp(C/TS_c) \quad (1)$$

with τ_0 and C nearly constant. The fundamental outcome as given by equation (1) appears to be preserved by modern treatments¹¹ of entropy-based theories. Indeed, the extent to which results of experimental studies performed by varying temperature¹² and pressure^{13,14} agree with equation (1) is surprising; results of computer simulations performed at different temperatures and densities^{15,16} on model glass formers similarly agree with equation (1). This agreement goes beyond what would be expected from a naïve model, as any entropy theory seems to be. It is now our aim to show that chemical vitrification shares with physical vitrification the same fundamental interpretation.

An entropy-based picture would qualitatively apply to polymerization reactions if changes in physical and chemical properties occurred on a timescale larger than that of thermodynamic equilibrium, so that the system's evolution could be seen as a succession of quasi-equilibrium states. In the entropy view, a dearth of available configurations expresses itself in the cooperative motion of the system, in such a way that the lower the configurational entropy, the greater the size of the cooperatively rearranging regions. As reaction proceeds, an increasing number of loose van der Waals bonds are replaced by stiffer covalent bonds that are expected to impose configurational restrictions and force the molecules to move cooperatively over an increasing length scale. On the other hand, experiments show a continuous relaxational slow-down occurring throughout the reaction, and suggest naturally a correlation with loss of configurations. Although intuitive, such a conjecture has not been proven so far but only taken for granted^{5,17}. A direct proof needs parallel information on how S_c changes throughout the reaction. Unfortunately, the character of the process prevents an experimental determination of S_c as the difference of total and vibrational entropy (as can be performed in stable systems¹²), and requires an alternative method to be found. A concrete possibility is shown below.

It is not obvious how to find a relationship that links a decrease in configurational entropy to the advancement of reaction, by a measurable chemical parameter. Here we identify step polymerizations¹⁸ as a convenient class of processes where configurational entropy can be related to chemical conversion, $\alpha(t)$, which measures the extent of reaction by the fraction (α) of functional groups reacted at elapsed time t since the beginning. Owing to the random nature of the growth mechanism in step polymerization, the degree of polymerization increases steadily throughout the reaction, but neither long chains nor large networks are formed on average until very high conversions are reached, even though the monomer is rapidly consumed early in the reaction. In such a situation, at any fixed temperature, one can expect monomers linked with each other to behave approximately as a single cooperative unit. As they are interlocked segments, a change in configuration of any one of them causes a change in configuration of the others, and it is realistic to expect that the average number of different configurations accessible to one cooperative region remains the same. This argument can be formalized as $W_c(\alpha) = W_c(0)^{1/x_n}$, where x_n is the number-average degree of polymerization (that is, the average number of monomers per molecule), $W_c(\alpha)$ is the number of configurations available to the macroscopic system at conversion α , and $W_c(0)$ the number of configurations at $\alpha = 0$. Accordingly, for the configurational entropy, which is given by $S_c \equiv k_B \ln W_c$ with k_B Boltzmann's constant, one has $S_c(\alpha) = S_c(0)/x_n$. However crude this picture may appear, we think that it is not cruder than the entropy theory itself: the most severe test would be a comparison with experimental results, once the expression above is inserted into equation (1). On this basis we propose a very simple relation, whose excellent agreement with experiments (shown later) demonstrates the validity of our reasoning.

For a mixture of monomers bearing functional groups of type A and monomers with functional groups of type B (where A groups can react only with B groups and vice versa), provided that no rings are formed, x_n can always be obtained by elementary stoichiometry¹⁹ as $x_n(\alpha_A) = 1/(1 - \langle f_A \rangle \alpha_A)$, where α_A is the fraction of A groups that have reacted, and $\langle f_A \rangle$ is the average number per monomer of A groups (a similar relation holds with B replacing A; hence the subscript will be omitted from now on for simplicity of notation, by keeping in mind that $\langle f \rangle$ and α must refer to the same kind of functional group). Thereby, one has $S_c(\alpha) = S_c(0)(1 - \langle f \rangle \alpha)$, which translates equation (1) into

$$\tau = \tau_0 \exp \left(\frac{B(T)}{1 - \langle f \rangle \alpha} \right) \quad (2)$$

with τ_0 approximately independent of both T and α , and $B(T) = [C/TS_c(0)]$ only dependent on temperature, on the assumption that the variation of C with α is negligible. Support for the last assumption can be inferred from refs 10 and 11. Equation (2) lends itself to stringent experimental tests, owing to its very strong predictions as follows. First, the condition of 'ideal' glass transition (where τ tends to infinity owing to S_c becoming zero) is that α equals the theoretical value $1/\langle f \rangle \equiv \alpha_0^{\text{theor}}$; second, because the average functionality $\langle f \rangle$ may be controlled by varying the molar ratio of reagents, α_0^{theor} may also be varied in this way irrespective of any reaction details; and third, for a given reaction, the same α_0^{theor} is expected at different T .

With this in mind, we test equation (2) for many different systems, using relaxation times measured by two different spectroscopic techniques, both sensitive to structural dynamics. For each reaction, carried out isothermally, experimental data of τ as function of conversion are fitted by the form $\tau = \tau_0 \exp[B/(1 - \alpha/\alpha_0)]$, with τ_0 , B and α_0 as free parameters. The fit parameters are listed in Table 1 for the set of the 17 polymerizations studied. When plotted against $(1 - \alpha/\alpha_0)^{-1}$, $\log_{10} \tau$ should display linear behaviour. This data representation is shown in Fig. 1, which emphasizes the remarkable agreement (over the whole range investigated) with the functional dependence on α expected from equation (2).

More importantly, the results in Table 1 meet our expectations for the values of the fitting parameters, including their variation with T . In all cases examined, the divergence of τ and therefore the graphic linearization in Fig. 1 occurs, within quite acceptable experimental error, strictly around the ideal values $\alpha_0^{\text{theor}} = 1.00$, 0.88 and 0.75 that are expected for our mixtures solely on the basis of their functionality and molar ratios, independently of T . This result constitutes strong evidence in favour of the idea behind equation (2). Further, the T dependence of the pre-exponential factor τ_0 is expected to be negligible, and the T dependence of the B parameter is expected to be related to that of $S_c(0)$ and C . One can reasonably assume that the variation with T of $S_c(0)$ overwhelms the variation of C , related to a rearranging unit's activation free energy. From $S_c(0, T) = \int_{T_0}^T (\Delta C_p/T') dT'$ (with ΔC_p the configurational heat capacity, and T_0 the ideal glass transition temperature²⁰ of the unreacted mixture), on condition that the approximation $\Delta C_p \propto$

Table 1 Fitting parameters

System	T (°C)	α_0^*	$\log[\tau_0^{-1} (\text{s}^{-1})]$	B	Techn.†	Refs‡
EPON828/EDA 1:1	25	1.00 ± 0.03	9.4 ± 0.1	2.9 ± 0.1	DS	4, PW
EPON828/EDA 1:1	25	1.00 ± 0.03	7.8 ± 0.3	2.7 ± 0.3	PCS	PW
EPON828/BAM 1:1	25	1.01 ± 0.02	9.5 ± 0.1	2.3 ± 0.1	DS	7, PW
EPON828/CHA 1:1	27	1.03 ± 0.03	12.7 ± 0.4	8.8 ± 0.5	DS	9, PW
EPON828/CHA 1:1	41	1.05 ± 0.03	12.3 ± 0.2	7.0 ± 0.3	DS	5
DGEBA348/DETA 4:3	25	0.87 ± 0.02	7.9 ± 0.2	3.1 ± 0.2	PCS	PW
DGEBA348/DETA 4:3	30	0.90 ± 0.02	7.8 ± 0.2	2.8 ± 0.2	PCS	PW
EPONX22/DDM 2:1	55	0.75 ± 0.02	8.8 ± 0.3	2.5 ± 0.2	DS	2
EPONX22/DDM 2:1	60	0.77 ± 0.03	8.4 ± 0.5	2.6 ± 0.4	DS	2
EPONX22/DDM 2:1	68	0.79 ± 0.03	7.9 ± 0.5	2.2 ± 0.4	DS	2
EPONX22/DDM 2:1	75	0.77 ± 0.03	8.0 ± 0.3	1.3 ± 0.3	DS	2
EPONX22/DDM 2:1	82	0.78 ± 0.03	8.5 ± 0.3	2.0 ± 0.4	DS	2
EPONX22/DDM 2:1	90	0.80 ± 0.03	8.7 ± 0.7	1.5 ± 0.4	DS	2
EPONX22/DDM 2:1	98	0.79 ± 0.03	8.1 ± 0.9	1.0 ± 0.4	DS	2
EPON828/EDA 2:1	23	0.76 ± 0.03	8.2 ± 0.7	5.3 ± 0.9	PCS	PW
EPON828/EDA 2:1	25	0.80 ± 0.03	8.3 ± 0.4	4.3 ± 0.4	PCS	PW
EPON828/EDA 2:1	32.1	0.75 ± 0.01	8.3(fixed)	5.5 ± 0.2	PCS	PW

Results of fitting the equation $\tau = \tau_0 \exp[B/(1 - \alpha/\alpha_0)]$ to experimental data for isothermal step polymerizations at temperature T . The values α_0 obtained experimentally must be compared with the values $\alpha_0^{\text{theor}} = 1/\langle f \rangle$ predicted theoretically, which are: $\alpha_0^{\text{theor}} = 1.00$ (for 1:1 mixtures), $\alpha_0^{\text{theor}} = 0.88$ (for 4:3 mixtures) and $\alpha_0^{\text{theor}} = 0.75$ (for 2:1 mixtures). For a given reaction, the same α_0 is expected for different T . Within experimental limits, equation (2) proves to be successful in (1) linear and network polymerizations; (2) different molar ratios of the same reagents (for example, both stoichiometric and non-stoichiometric balance of mutually reactive functional groups is realized in the EPON828/EDA mixture); and (3) different curing temperatures, for the same reacting mixture. The findings are confirmed by different physical observables, measured by DS and PCS, probing structural dynamics. EDA, BAM and so on are defined in Fig. 1 legend.

*An uncertainty of the order of 2–3% is due to errors inherent in the experimental determination of chemical conversion.

†Techniques are: DS, dielectric spectroscopy; PCS, depolarized photon-correlation spectroscopy.

‡PW, present work. Data from the literature are also included.

$1/T$ holds—as several experimental data indicate¹²— $B(T)$ takes the form $B(T) = D/(T - T_0)$ with D a constant, and yields a Vogel–Fulcher–Tamman²¹ form to equation (2) at any fixed conversion. Trends for τ_0 and B resulting from the fit are clearly consistent with these expectations (Table 1, and Fig. 1b inset). We notice that a decreasing behaviour with α of the ‘kinetic fragility’, a parameter

quantified by the index K_{VFT} in the Vogel–Fulcher–Tamman equation written as $\tau = \tau_0 \exp(1/[K_{VFT}(T/T_0 - 1)])$, is implicit in equation (2) using $B(T) = D/(T - T_0)$. This behaviour is consistent with the presence of an increasing fraction of highly directional interactions between molecules, as typical of ‘strong’ systems²². The few experiments existing confirm that fully reacted mixtures are less ‘fragile’ than the unreacted ones⁸.

Replacing phenomenological $\tau(\alpha)$ equations with a physically based equation should lead to benefits for materials processing. We now consider what we can learn about general glass-formation from our findings. The validity of equation (2) has two implications. First, it validates the proposed relationship between the number of chemical bonds created and the associated reduction in configurational entropy; and second, it provides convincing evidence that the connection between relaxation behaviour and configurational entropy embodied in equation (1) also survives in chemical glass formers during the irreversible formation of chemical bonds. We conclude that configurational entropy acts as a thermodynamic controller of the slow dynamics in chemically vitrifying systems, just like it does in glass-forming liquids under cooling or compression.

Energy landscape (the hypersurface described by potential energy as a function of all the coordinates of a system)^{15,16,23–25} offers a convenient microscopic perspective for interpreting the key role of configurational entropy in the dynamics of physical and chemical glasses. The formation of new covalent bonds can be connected to events that originate forbidden paths randomly distributed in configuration space, and finally generate forbidden regions. The number of basins that remain accessible to the system is reduced, and configurational entropy, which measures this number, becomes lower. Decreasing the volume in a liquid also causes the distribution of basins to change^{16,26}, by producing a numerical reduction of accessible minima, very similar to the effect of cooling. Such similarity strongly suggests that S_c may become the unifying concept in the physics of systems characterized by slow dynamics, well beyond the restricted world of structural glasses where this concept was introduced.

We consider that an entropy-based treatment holds the promise of understanding several phenomena—in systems such as physical gels, microemulsions and disordered systems—where the dynamics are phenomenologically similar to vitrification. One phenomenon of particular interest is thermoreversible gelation in physically associating polymers, whose nature is not yet established, and where geometrical percolation is found not to guarantee solid-like behaviour²⁷, rather like network polymerizations where gel formation does not correspond to structural arrest. In analogy with the ideas developed here, the diffusion in associating polymers might be connected to the average degree of association.

Our results show not only that vitrification dynamics are driven by configurational entropy, but that the dynamics-to-thermodynamics correlation, $\tau(S_c)$, appearing in the form of equation (1) along physical vitrification paths (by varying T or V) is preserved in the same functional form along chemical vitrification paths (by varying α). As processes of a chemical nature result in modifications of the interaction potential between molecules so strong as to change the molecular architecture, our findings suggest a nontrivial universality in the relationship of liquid structure to dynamics in dense liquids, independent of the interaction potential and system specifications. The existence of such a universal property is by no means obvious, as emerges from great diversity at a microscopic level. The energy-landscape approach may be the formalism that best unravels the connection between S_c and the universal features characterizing slow dynamics in different materials. At each α and V value there is a corresponding fixed landscape, and the temperature rules the manner in which the landscape is explored. A thermodynamics-to-dynamics correlation implies the existence of topological constraints between basin minima and saddle points that allow

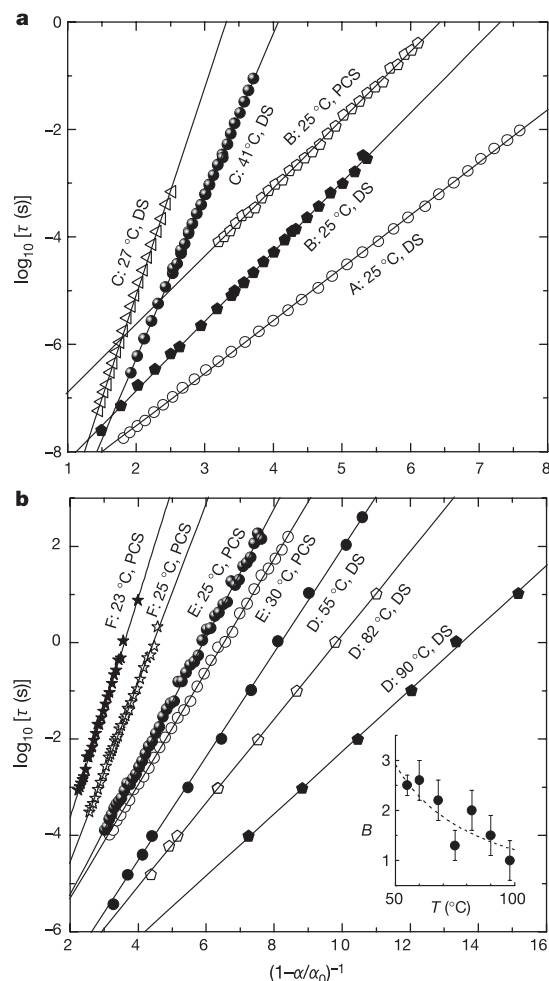


Figure 1 Dependence of the structural relaxation time, τ , on the conversion of epoxy groups, α , for several isothermal step polymerizations of epoxy-amine systems, where reaction proceeds by polyaddition. Three epoxy prepolymers (diglycidyl-ether of bisphenol A) are used, with functionality $f = 2$ and different epoxy equivalent weight: EPON828, EPONX22 and DGEBA348. Ring formation is negligible in the curing of these systems with amines²⁹. Either linear or network polymers are obtained, depending upon the functionality of the curing agent. Results shown refer both to reactions with bifunctional amines (*n*-butylamine (BAM) and cyclohexylamine (CHA)), yielding linear chain polymers, and to reactions with multifunctional amines (ethylenediamine (EDA) and 4,4'-diaminodiphenylmethane (DDM) with $f = 4$, and diethylenetriamine (DETA) with $f = 5$), yielding network polymers. Relaxation times are obtained from dielectric spectroscopy (DS) and depolarized photon-correlation spectroscopy (PCS) as $\tau = 1/2\pi\nu_{\max}$, where $\nu_{\max}$ is the frequency at which the imaginary part of the measured susceptibility has a maximum. The conversion is measured by calorimetry (refs 3–7 and present work), except for the EPONX22/DDM 2:1 system where it is obtained by Fourier-transform infrared spectroscopy². Data are plotted as $\log_{10} \tau$ versus $(1 - \alpha/\alpha_0)^{-1}$, with α_0 the conversion at which τ tends to diverge (see values in Table 1). **a**, Experimental results for some reactions with $\alpha_0^{\text{theor}} = 1.00$. The systems are indicated on the plot as: A, EPON828/BAM 1:1; B, EPON828/EDA 1:1; and C, EPON828/CHA 1:1. **b**, Results for some reactions with $\alpha_0^{\text{theor}} = 0.75$ (EPONX22/DDM 2:1 and EPON828/EDA 2:1) and $\alpha_0^{\text{theor}} = 0.88$ (DGEBA348/DETA 4:3). The systems are indicated on the plot as: D, EPONX22/DDM 2:1; E, DGEBA348/DETA 4:3; F, EPON828/EDA 2:1. The inset shows the temperature dependence of the parameter B in equation (2), for the EPONX22/DDM 2:1 system. The dashed line is a fit with $B(T) = D/(T - T_0)$.

the system to change basin. Our results suggest that comparable constraints must characterize those regions of configuration space explored at different temperatures in a fixed landscape, and regions explored at a given temperature in a family of energy landscapes resulting from either covalent-bond formation or volume reduction. Thus, the general applicability of equation (1) evidently rests on universal properties in the organization of energy landscapes, and microscopic diversity must be absorbed into statistical properties, possibly governed by limit theorems. Manifestations of that might be looked for in simulation studies.

Indeed, a glass can be obtained both as a result only of compression, and as a result only of reaction, as the underlying evolution of configurational restrictions turns out to be similar. Nevertheless, the mechanism of disappearance of accessible minima is different in the two cases. Whereas a volume change involves a simultaneous modification of the interactions in all parts of the system and alters everywhere its configuration space, the formation of any new covalent bond does not entail a long-range modification of the energy landscape. The resulting landscapes may be topologically different. In this connection, we note that kinetic fragility (which has a thermodynamic^{16,28} measure in the rapidity of change with T of configurational entropy) is likely to reflect such difference; indeed, the kinetic fragility is expected from equation (2) to become lower with increasing α , whereas it is found to become higher with decreasing volume¹⁶. □

Received 7 August; accepted 23 October 2002; doi:10.1038/nature01261.

1. Debenedetti, P. G. *Metastable Liquids. Concepts and Principles* (Princeton Univ. Press, Princeton, 1996).
2. Deng, Y. & Martin, C. Analysis of the cure-dependent dielectric relaxation behavior of an epoxy resin. *J. Polym. Sci. B* **32**, 2115–2125 (1994).
3. Cassettari, M., Salvetti, G., Tombari, E., Veronesi, S. & Johari, G. P. Dielectrics and thermodynamics of a macromolecule's growth. *J. Non-Cryst. Solids* **172–174**, 554–561 (1994).
4. Casalini, R., Corezzi, S., Fioretto, D., Livi, A. & Rolla, P. A. Unified dielectric description of the dynamics of polymeric systems undergoing either thermal or chemical vitrification. *Chem. Phys. Lett.* **258**, 470–476 (1996).
5. Tombari, E., Ferrari, C., Salvetti, G. & Johari, G. P. Molecular dynamics during linear chain polymerization from real-time dielectric spectrometry and calorimetry. *J. Phys. Condens. Matter* **9**, 7017–7037 (1997).
6. Johari, G. P., Ferrari, C., Tombari, E. & Salvetti, G. Temperature modulation effects on a material's properties: Thermodynamics and dielectric relaxation during polymerization. *J. Chem. Phys.* **110**, 11592–11598 (1999).
7. Gallone, G., Capaccioli, S., Levita, G., Rolla, P. A. & Corezzi, S. Dielectric analysis of the linear polymerization of an epoxy resin. *Polym. Int.* **50**, 545–551 (2001).
8. Parthun, M. G. & Johari, G. P. Dielectric spectroscopy of a polymerizing liquid and the evolution of molecular dynamics with increase in the number of covalent bonds. *J. Chem. Phys.* **103**, 440–450 (1995).
9. Johari, G. P., Ferrari, C., Salvetti, G. & Tombari, E. Physico-chemical aspects of dielectric and thermodynamic changes during high-temperature polymerization and their technical use. *Phys. Chem. Chem. Phys.* **1**, 2997–3005 (1999).
10. Adam, G. & Gibbs, J. H. On the temperature dependence of cooperative relaxation properties in glass-forming liquids. *J. Chem. Phys.* **43**, 139–146 (1965).
11. Xia, X. & Wolynes, P. G. Fragilities of liquids predicted from the random first order transition theory of glasses. *Proc. Natl Acad. Sci. USA* **97**, 2990–2994 (2000).
12. Richert, R. & Angell, C. A. Dynamics of glassforming liquids. V: On the link between molecular dynamics and configurational entropy. *J. Chem. Phys.* **108**, 9016–9026 (1998).
13. Casalini, R., Capaccioli, S., Lucchesi, M., Rolla, P. A. & Corezzi, S. Pressure dependence of structural relaxation time in terms of the Adam-Gibbs model. *Phys. Rev. E* **63**, 031207 (2001).
14. Casalini, R. *et al.* Effect of pressure on the dynamics of glass formers. *Phys. Rev. E* **64**, 041504 (2001).
15. Scala, A., Starr, F. W., La Nave, E., Sciortino, F. & Stanley, H. E. Configurational entropy and diffusivity of supercooled water. *Nature* **406**, 166–169 (2000).
16. Sastry, S. The relationship between fragility, configurational entropy and the potential energy landscape of glass-forming liquids. *Nature* **409**, 164–167 (2001).
17. Matsuoka, S., Quan, X., Bair, H. E. & Boyle, D. J. A model for the curing reaction of epoxy resins. *Macromolecules* **22**, 4093–4098 (1989).
18. Young, R. J. & Lovell, P. A. *Introduction to Polymers* (Chapman and Hall, New York, 1991).
19. Macosko, C. W. & Miller, D. R. A new derivation of average molecular weights of nonlinear polymers. *Macromolecules* **9**, 199–206 (1976).
20. Kauzmann, W. The nature of the glassy state and the behavior of liquids at low temperatures. *Chem. Rev.* **43**, 219–256 (1948).
21. Vogel, H. Temperature dependence of viscosity of melts. *Phys. Z.* **22**, 645–646 (1921).
22. Angell, C. A. Relaxation in liquids, polymers and plastic crystals—strong/fragile patterns and problems. *J. Non-Cryst. Solids* **131–133**, 13–31 (1991).
23. Stillinger, F. H. A topographic view of supercooled liquids and glass formation. *Science* **267**, 1935–1939 (1995).
24. Sastry, S., Debenedetti, P. G. & Stillinger, F. H. Signatures of distinct dynamical regimes in the energy landscape of a glass-forming liquid. *Nature* **393**, 554–557 (1998).
25. Debenedetti, P. G. & Stillinger, F. H. Supercooled liquids and the glass transition. *Nature* **410**, 259–267 (2001).

26. La Nave, E., Mossa, S. & Sciortino, F. Potential energy landscape equation of state. *Phys. Rev. Lett.* **88**, 225701 (2002).
27. Kumar, S. K. & Douglas, J. F. Gelation in physically associating polymer solutions. *Phys. Rev. Lett.* **87**, 188301 (2001).
28. Martinez, L.-M. & Angell, C. A. A thermodynamic connection to the fragility of glass-forming liquids. *Nature* **410**, 663–667 (2001).
29. Luňák, S. & Dušek, K. Curing of epoxy resins. II. Curing of bisphenol A diglycidyl ether with diamines. *J. Polym. Sci. Polym. Symp. Edn* **53**, 45–55 (1975).

Acknowledgements We thank G. Gallone and S. Capaccioli for providing dielectric measurements on EPON828/EDA 1:1 and EPON828/BAM 1:1 systems, and L. Comez, P. Grigolini, S. Mossa, G. Ruocco, A. Scala and G. Socino for comments on the manuscript. We particularly thank F. Sciortino for assistance.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to S.C. (e-mail: Silvia.Corezzi@fisica.unipg.it).

Decreased stability of methane hydrates in marine sediments owing to phase-boundary roughness

W. T. Wood*, J. F. Gettrust*, N. R. Chapman†, G. D. Spence‡ & R. D. Hyndman‡

* US Naval Research Laboratory, Stennis Space Center, Mississippi 39529, USA

† University of Victoria, Victoria, British Columbia, Canada V8W 3P6

‡ Pacific Geosciences Center, Sydney, British Columbia, Canada, V8L 4B2

Below water depths of about 300 metres, pressure and temperature conditions cause methane to form ice-like crystals of methane hydrate¹. Marine deposits of methane hydrate are estimated to be large, amassing about 10,000 gigatonnes of carbon², and are thought to be important to global change^{3,4} and seafloor stability^{5,6}, as well as representing a potentially exploitable energy resource⁷. The extent of these deposits can usually be inferred from seismic imaging, in which the base of the methane hydrate stability zone is frequently identifiable as a smooth reflector that runs parallel to the sea floor. Here, using high-resolution seismic sections of seafloor sediments in the Cascadia margin off the coast of Vancouver Island, Canada, we observe lateral variations in the base of the hydrate stability zone, including gas-rich vertical intrusions into the hydrate stability zone. We suggest that these vertical intrusions are associated with upward flow of warmer fluids. Therefore, where seafloor fluid expulsion and methane hydrate deposits coincide, the base of the hydrate stability zone might exhibit significant roughness and increased surface area. Increased area implies that significantly more methane hydrate lies close to being unstable and hence closer to dissociation in the event of a lowering of pressure due to sea-level fall.

The data we present were acquired with a unique seismic system (deep-towed acoustic/geophysics system; DTAGS⁸) that can be towed close to the sea floor, resulting in a reduced Fresnel zone (lateral area over which a reflected signal is averaged) and increased lateral resolution. Also, because DTAGS is operated at 250–650 Hz, rather than 10–80 Hz, the resulting vertical and lateral resolutions are almost an order of magnitude greater than conventional seismic systems⁹. Deployment of DTAGS on the accretionary prism of the Cascadia margin, a region known for a strong bottom-simulating reflector (BSR), and widespread evidence of fluid flux and methane hydrate^{10–12}, resulted in some intriguing observations that are

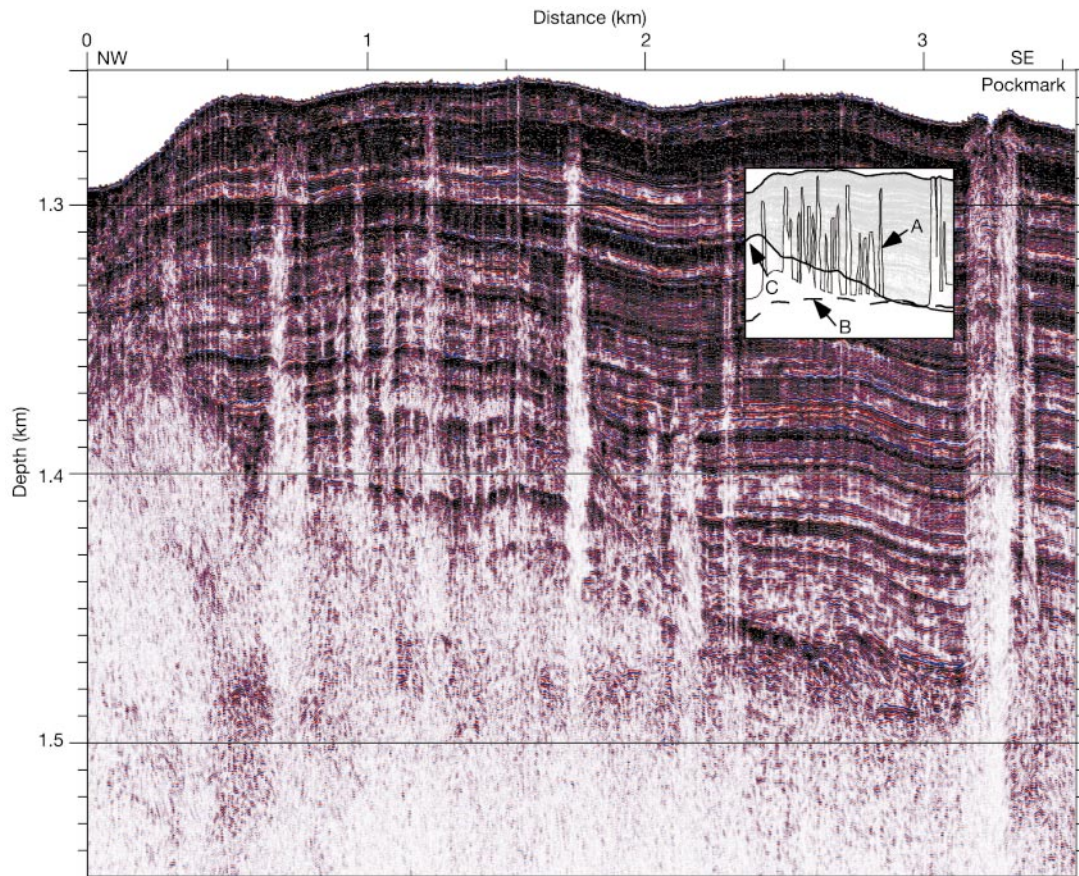


Figure 1 High-resolution seismic images of the hydrate stability zone (upper 220 m of sediment) on the Cascadia margin about halfway between Ocean Drilling Program sites 889 and 890 (ref. 10). Vertically oriented gas accumulations (curve A, inset) scatter and absorb the incoming energy, resulting in little acoustic energy returning upward to be recorded. Gas accumulations conformable with short pieces of near-horizontal strata

(curve B, inset) within the highly deformed accreted sediments (below curve C, inset) greatly enhance the energy coherently reflected upward. The presence of gas within the hydrate stability zone suggests that the zone might have been perturbed by warm fluids, occasionally resulting in a seafloor pockmark. These perturbations significantly increase the surface area and subsequently the volatility of the methane hydrate reservoir.

apparent in Fig. 1. The sediment column here exhibits vertical zones of little to no coherently reflected seismic energy (wipeouts), slightly wider at the bottom than at the top (Fig. 1 inset, fine curve A), occasionally intersecting the sea floor. At least some of the intersections are associated with pockmarks (Fig. 1). In addition, instead

of the strong BSR seen in conventional seismic data throughout this area¹¹, we see a diffuse cloud of short reflection segments becoming more intense when proximal to the wipeouts (Fig. 1 inset, curve B). Because BSRs are caused by the coherent reflectivity of methane gas accumulation at (or just below¹³) the depth of the base of gas

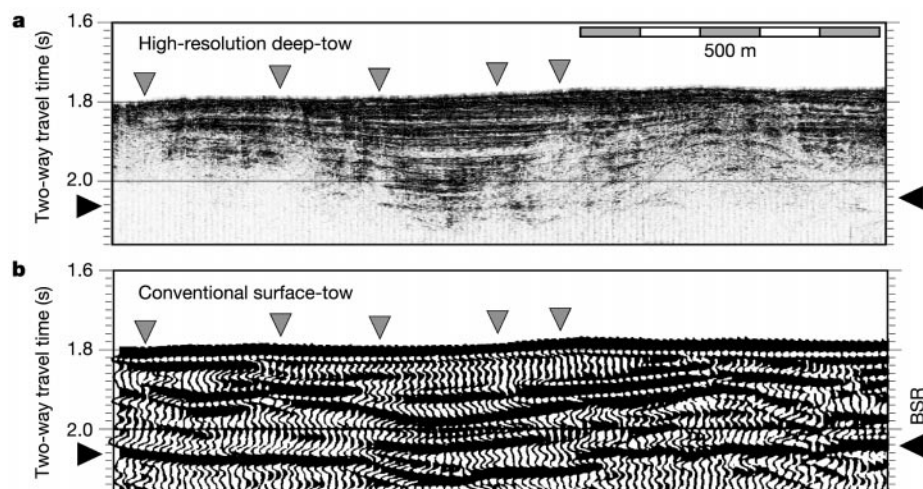


Figure 2 Two views of the same portion of the methane hydrate stability zone. Areas of reduced amplitude due to fine-scale flux conduits (grey arrowheads) are apparent in the

DTAGS data (a) but not in the airgun data (b, line 89-08 from ref. 10), whereas a coherent BSR (black arrowheads) is apparent in the airgun data but not the DTAGS data.

hydrate stability (BGHS), and are sensitive to fluid and heat flux¹, the wipeouts and perturbations imaged here are probably indicators of fine-scale fluid flux through the upper sea floor.

The level of detail in these images can be appreciated better where both DTAGS and conventional data are exactly coincident and displayed at the same scale, as in Fig. 2. The higher frequencies and smaller Fresnel zone of DTAGS allows the identification of the narrow wipeouts (grey arrowheads in Fig. 2) but do not have the penetration of conventional systems, especially in the more indurated accretionary wedge sediments that contain the BSR (black arrowheads in Fig. 2). It is also significant that the shorter wavelengths associated with higher frequencies require increased accuracy in source and receiver geometry¹², which is extremely difficult to achieve in any marine experiment. The optimal DTAGS images displayed here require an increased effort in data processing over that of the conventional seismic image.

Even with the higher-resolution images, the exact geologic cause of the wipeouts is ambiguous; we are certain only that the seismic energy is not coherently reflected back to the receiver array. The energy can be absorbed (for example, by gas) or scattered, perhaps by sub-metre-scale unevenly distributed accumulations of gas, methane hydrate or calcium carbonate (CaCO₃), or by stratal disruption from intense fluid flux. However, the seafloor pockmarks and carbonate mineralization associated with some wipeouts are clear indicators (if only qualitative) of upward fluid flux and, by inference, heat flux.

Several factors lead us to propose that the wipeouts are caused by gas, the outline of which coincides with columnar perturbations in the BGHS. First, the wipeouts are strikingly similar in size and character to 'gas chimneys'^{214,15}, vertically oriented accumulations of gas which have been seen in conventional seismic data from several parts of the world, and in water depths too shallow to sustain

methane hydrate¹⁶. In addition, localized heat flux mediated by advecting fluid implies that the pressure–temperature boundary defining the BGHS is perturbed upwards from the regional average at least mildly, and perhaps quite severely, especially where the wipeout is coincident with a seafloor pockmark, creating a region within which methane hydrate dissociates and releases gas. Because methane hydrate at depth is warmer and more easily dissociated than methane hydrate near the sea floor, we expect the perturbation to broaden with depth, which is consistent with the observed outline of many of the wipeouts.

This interpretation of the wipeouts as gas chimneys is also consistent with preliminary finite element modelling (SUTRA¹⁷), using actual measurements of temperature, pressure and physical properties from this area¹⁰, and measured properties of methane hydrate stability¹⁸. The wipeout shown in Fig. 3 is an expanded view of the wipeout located at 2.3 km in Fig. 1, and was modelled in three dimensions assuming radial symmetry about a vertical zone with a permeability three orders of magnitude greater than the surrounding sediments (see Methods for details). Every modelling attempt resulted in a flared base to the stability boundary (dashed line in Fig. 3), which is consistent with the outline of this and most other wipeouts. The absolute height and width of the perturbation depend on conduit permeability and on temporal functions of fluid flux rates and temperature, none of which are well constrained on the scale of metres to tens of metres, but the results demonstrate

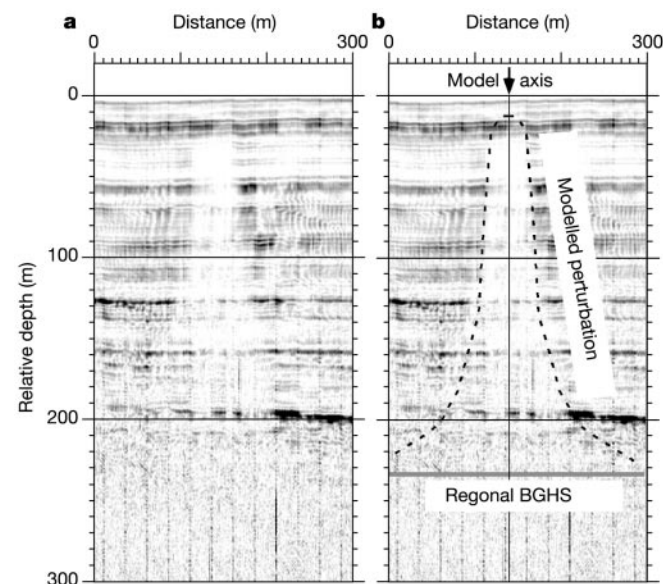


Figure 3 Finite element modelling results for heat and fluid flux of the wipeout at 2.3 km in Fig. 1. **a**, This wipeout is a contrast in seismic reflection amplitude emphasized here by displaying the instantaneous amplitude of the signal (light and dark shades are low and high amplitudes, respectively) and using relatively low gain. **b**, Finite element modelling of heat and fluid flux was performed assuming radial symmetry about a vertical flux conduit axially located within the wipeout. After many tens of thousands of years of model time the temperature field stabilized and the resulting perturbation (dashed line) of the regional stability boundary (grey horizontal line) was calculated. The downward flare of the perturbation is consistent with the boundary of this and most other wipeouts.

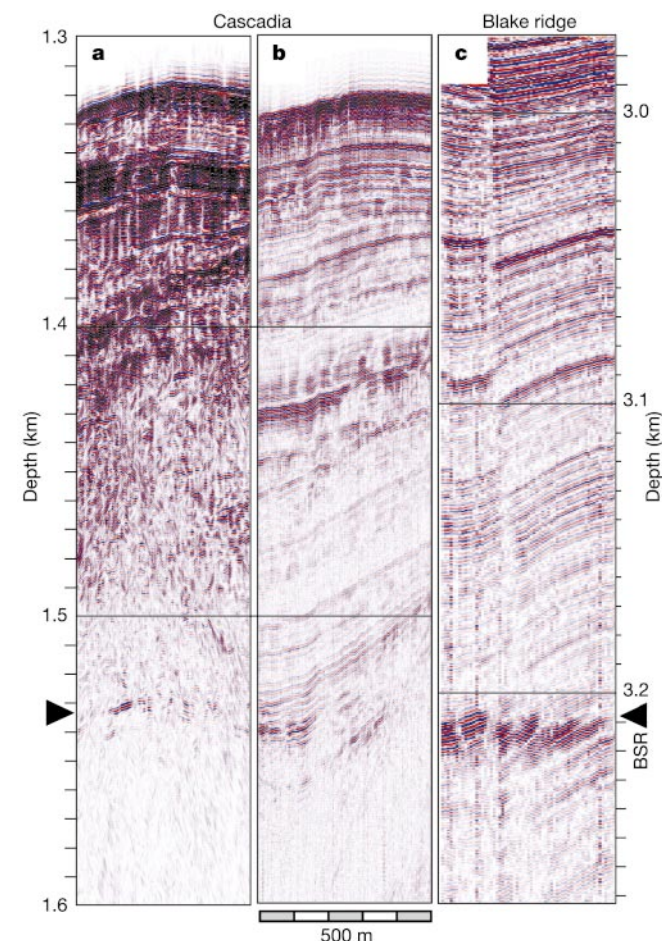


Figure 4 A comparison of BSRs. Where gas accumulations are small relative to the ensconifying wavelength and the Fresnel zone (**a**), the BSR appears as a single reflection. Where the gas accumulation is larger, and strata are resolvable (**b**, **c**), the BSR is a reflective zone of enhanced reflectivity conformable with the strata.

that, in areas of fluid and heat flux, methane gas can exist well inside the regional hydrate stability zone.

Thus in our interpretation the chimneys represent a significantly increased surface area of the BGHS, increasing the responsiveness of the methane hydrate reservoir to methane release or sequestration in the event of sea-level fall or rise, respectively. When plotted with no vertical exaggeration, the interpreted BGHS (curve A in Fig. 1 inset) is about 1.64 times as long as the expected, or unperturbed, BGHS (curve B in Fig. 1 inset). If we assume that the distribution of chimneys in the surveyed area is the same across the section in Fig. 1 as it is within the section, we find anything from a 0.5-fold to a 3.0-fold increase in the surface area of the phase boundary (depending on the interpretation of the wipeout boundary) and therefore the volatility of the sequestered hydrate. The chimneys are also important from a purely practical standpoint, representing remotely detectable fluid flux conduits that are probably associated with increased hydrate concentrations close to the stability boundary, and are of primary importance in any physical, chemical, biological or industrial study of methane hydrates^{1,7,19}.

The other phenomenon observed in the DTAGS data also relates to the roughness of the BGHS, but at a much smaller scale. The data also reveal BSRs with a significantly different character from that of the strong, reversed-polarity, strata-crossing reflection frequently observed in surface-tow seismic data such as those in Fig. 2. Published modelling of a wide range of data including conventional and DTAGS data shows that much of the reduced amplitude can be explained by a vertical velocity gradient that is gentle (less reflective) for high-frequency data and sharper (more reflective) for low-frequency data¹².

The new images do not refute the gradient model, but they suggest that lateral variability is likely to be a significant factor in the appearance of the BSR. The data indicate that anywhere the strata containing the BSR are resolvable, the BSR appears only as an upper extent of reflective sediments below and is not an unconformable reflection. Only when the spatial scale of the sedimentary structure falls below the lateral resolution of the seismic system in use does the BSR seem to cut across local strata. This is consistent with a BSR caused by localized accumulations of gas preferentially occupying more porous strata, thereby merely enhancing the existing reflectivity instead of creating an additional reflector. In contrast to the much larger-scale chimneys, the BGHS need not be perturbed to exhibit this character.

An example of this lateral variability is shown in Fig. 4b, where the BSR is more parallel to surrounding sediments at the left than to those at the right, giving the BSR a more continuous appearance on the left and a shingled appearance on the right. A very similar transition is seen in Fig. 4c in DTAGS data acquired on the Atlantic side of the continent over a large sediment drift, the Blake ridge, near Ocean Drilling Program site 997, another region known for a prominent BSR and hydrate accumulations^{20,21}. Again, the BSR undergoes a transition from parallel to shingled from left to right, although the section from the Blake ridge exhibits a sharper distinction between the gas-free sediments above and the gas-charged sediments below. Geologically and hydrologically, the two settings are quite different except that both areas exhibit conspicuous BSRs and are made up of fine-grained sediments typical of continental margins, where most methane hydrates are suspected to occur².

In areas of greater sediment deformation, such as the accreted sediments of the Cascadia margin, no sedimentary strata are resolvable, and the short coherent segments of reflections that we see in the DTAGS data (the diffuse cloud in Fig. 1, or the intermittent single reflector in Fig. 4a) might well be laterally oriented assemblages of sub-metre-scale accumulations that appear coherent only when averaged over the first Fresnel zone of the ensonifying seismic wave. This is analogous to the individually resolvable segments in Fig. 4c seeming coherent in lower-frequency surface-tow data²¹.

Last, the BSR in the DTAGS data is most conspicuous immediately adjacent to chimneys (Fig. 1) and to some extent the faults (Fig. 4). Note the thicker region of intense reflections near the bases of the more prominent chimneys in Fig. 1. This is consistent with fluid flux focused at the chimneys enriching the surrounding sediment with methane. Within the wipeout we expect the free gas to be oriented along the near-vertical flux conduits, effectively scattering the seismic signal; below the BGHS adjacent to the wipeout we expect near-horizontally oriented free gas, occupying pore space within preferentially permeable and porous strata, and enhancing their coherent seismic reflectivity (the BSR). Although there is no direct evidence, we would also expect an increased concentration of gas hydrate filling pores immediately adjacent to the wipeout above the regional BGHS, mimicking the increased free gas below the regional BGHS.

The high-resolution images of the methane hydrate stability zone and its lower boundary depicted here support a model of hydrate distribution strongly affected, if not dominated, by focused vertical fluid flux. This complicates the estimation of methane hydrate quantities obtained through extrapolation from a single drill hole, and indicates that of the estimated 10,000 gigatonnes of carbon sequestered in methane hydrates, a significantly greater portion might lie closer to the stability boundary than if it were evenly distributed throughout an unperturbed stability zone. □

Methods

Preliminary modelling of BGHS perturbation from advective heat flux was performed with the finite element program SUTRA¹⁶. Figure 3 shows the modelling result from one of the seismic wipeouts from the Cascadia margin. A three-dimensional result was obtained with a two-dimensional model by assuming radial symmetry about a vertical flux conduit. Each of the 30 × 30 elements used represents a region 10 m in depth by 10 m outwards from the conduit. The conduit was modelled by setting the hydraulic conductivity of the inner two columns to be three orders of magnitude greater than in the surrounding sediments.

Appropriate initial and boundary conditions as well as sediment and fluid properties used in the SUTRA modelling were based on published measurements made at Ocean Drilling Program sites 889 and 890 (ref. 10) only a few kilometres away from the modelled wipeout and in the same geologic setting. Fraction porosity (ϕ), pressure in MPa (P) and temperature in °C (T) were all given initial conditions that were linear functions of depth in metres, z (positive downwards), namely $\phi(z) = 0.63 - 0.0005z$, $P(z) = 13.0585 + 0.012054z$ and $T(z) = 2.6 + 0.0545z$. The pressure and temperature on the model boundaries, that is, the top of the model (sea floor), base of the model (at 300 m below the sea floor) and outer edge of the model (300 m radius from the conduit) were held fixed throughout the simulation at their initial values. The physical properties of the fluid and solid used in the model were respectively: compressibility, 4.9×10^{-10} and 0.0 Pa^{-1} ; specific heat, 4,182 and $4,840 \text{ J kg}^{-1} \text{ K}^{-1}$; thermal conductivity, 0.528 and $1.516 \text{ W m}^{-1} \text{ K}^{-1}$; and density, 1.035 and 2.65 g cm^{-3} . The value of the dispersion coefficient used in the model was 5 in all directions. In the conduit, vertical and lateral hydraulic conductivity are both $10^{-13} \text{ m s}^{-1}$; in the surrounding sediment, vertical and lateral hydraulic conductivities are 10^{-17} and $10^{-16} \text{ m s}^{-1}$, respectively.

The model was allowed to proceed to 10,000 years model time, when the change between subsequent 100-year time steps was negligible. The constant-pressure boundary at the base of the model allowed greater fluid and heat flux through the high-permeability conduit than through the surrounding sediment, thus perturbing the thermal profile. The equilibrium thermal profile provided by SUTRA was then used with the assumed pressure profile and a measured hydrate stability curve¹⁸ to determine the perturbation (dashed line in Fig. 3) of the regional BGHS (grey line in Fig. 3). Although several key modelling conditions (such as pressure and hydraulic conductivity) are poorly constrained, the characteristic shape of the thermal perturbation is, as expected, wider at the base than at the top. Although exhaustive testing has not yet been performed, every model run exhibited a downward flare in the BGHS, which is consistent with the observed profile of most of the wipeouts.

Received 20 May; accepted 28 October 2002; doi:10.1038/nature01263.

1. Paull, C. K. & Dillon, W. P. (eds) *Natural Gas Hydrates: Occurrence, Distribution, and Dynamics* (AGU Monograph no. 124, American Geophysical Union, Washington DC, 2001).
2. Kvenvolden, K. A. & Lorenson, T. D. in *Natural Gas Hydrates: Occurrence, Distribution, and Dynamics* (eds Paull, C. K. & Dillon, W. P.) 3–18 (AGU Monograph no. 124, American Geophysical Union, Washington DC, 2001).
3. Paull, C. K., Ussler, W. & Dillon, W. P. Is the extent of glaciation limited by marine gas-hydrates? *Geophys. Res. Lett.* **18**, 432–434 (1991).
4. Dickens, G. R., Castillo, M. M. & Walker, J. C. G. A blast of gas in the latest Paleocene: simulating first-order effects of massive dissociation of oceanic methane hydrate. *Geology* **25**, 259–262 (1997).
5. Paull, C. K., Buelow, W., Ussler, W. & Borowski, W. S. Increased continental-margin slumping frequency during sea-level lowstands above gas hydrate-bearing sediments. *Geology* **24**, 143–146 (1996).

6. Booth, J. S., Winters, W. J. & Dillon, W. P. Circumstantial evidence of gas hydrate and slope failure associations on US Atlantic continental margin. *Ann. N.Y. Acad. Sci.* **75**, 487–489 (1994).
7. Collett, T. S. & Kuuskraa, V. A. Hydrates contain vast store of world gas resources. *Oil Gas J.* 11 May, 90–95 (1998).
8. Gettrust, J. F., Ross, J. H. & Rowe, M. M. Development of a low frequency, deep-tow geacoustics system. *Sea Technol.* **32**, 23–32 (1991).
9. Wood, W. T. & Gettrust, J. F. in *Natural Gas Hydrates: Occurrence, Distribution, and Dynamics* (eds Paull, C. K. & Dillon, W. P.) 165–178 (AGU Monograph no. 124, American Geophysical Union, Washington DC, 2001).
10. Westbrook, G. K., Carson, B., Musgrave, R. J. & Suess, E. *Proc. ODP Init. Rep.* **146** (1994).
11. Hyndman, R. D. & Spence, G. D. A seismic study of methane hydrate marine bottom-simulating reflectors. *J. Geophys. Res.* **97**, 6683–6698 (1992).
12. Chapman, N. R. *et al.* High-resolution, deep-towed, multichannel seismic survey of deep-sea gas hydrates off western Canada. *Geophysics* **67**, 1038–1047 (2002).
13. Xu, W. & Ruppel, C. Predicting the occurrence, distribution, and evolution of methane gas hydrate in porous marine sediments. *J. Geophys. Res.* **104**, 5081–5095 (1999).
14. Heggland, R., Meldahl, P., de Groot, P. & Aminzadeh, F. Chimney cube unravels subsurface. *Am. Oil Gas Reporter* 78–83 (February 2000).
15. Meldahl, P., Heggland, R., Bril, B. & de Groot, P. Identifying faults and gas chimneys using multiattributes and neural networks. *Leading Edge* **20**, 474–482 (2001).
16. Hovland, M. & Judd, G. *Seabed Pockmarks and Seepages* 181 (Graham & Trotman, London, 1988).
17. Voss, C. I. A finite element simulation model for saturated and unsaturated, fluid-density-dependent ground-water flow with energy transport or chemically reactive single species solute transport (USGS Water Resources Investigations Report 84–4369, 1984).
18. Brown, K. M., Bangs, N. L., Froelich, P. N. & Kvenvolden, K. A. The nature, distribution and origin of gas hydrate in the Chile Triple Junction region. *Earth Planet. Sci. Lett.* **139**, 471–483 (1996).
19. Fisher, C. R. *et al.* Methane ice worms: *Hesiocaeca methanicola* colonizing fossil fuel reserves. *Naturwissenschaften* **87**, 184–187 (2000).
20. Paull, C. K., Matsumoto, R. & Wallace, P. *Proc. ODP Init. Rep.* **164** (1996).
21. Korenaga, J., Holbrook, W. S., Singh, S. C. & Minshull, T. A. Natural gas hydrates on the southeast U.S. margin: Constraints from full waveform and travel time inversions of wide angle seismic data. *J. Geophys. Res.* **102**, 15345–15365 (1997).

Acknowledgements We thank C. Voss for advice on finite element modelling.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to W.W. (e-mail: warren.wood@nrlssc.navy.mil).

Unconventional lift-generating mechanisms in free-flying butterflies

R. B. Srygley & A. L. R. Thomas

Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS, UK

Flying insects generate forces that are too large to be accounted for by conventional steady-state aerodynamics^{1,2}. To investigate these mechanisms of force generation, we trained red admiral butterflies, *Vanessa atalanta*, to fly freely to and from artificial flowers in a wind tunnel, and used high-resolution, smoke-wire flow visualizations to obtain qualitative, high-speed digital images of the air flow around their wings. The images show that free-flying butterflies use a variety of unconventional aerodynamic mechanisms to generate force: wake capture³, two different types of leading-edge vortex^{3–7}, active and inactive upstrokes⁸, in addition to the use of rotational mechanisms³ and the Weis-Fogh ‘clap-and-fling’ mechanism^{9–12}. Free-flying butterflies often used different aerodynamic mechanisms in successive strokes. There seems to be no one ‘key’ to insect flight, instead insects rely on a wide array of aerodynamic mechanisms to take off, manoeuvre, maintain steady flight, and for landing.

Recent experiments with tethered hawkmoths, and with large-scale mechanical flapping models, have successfully replicated some of the force-generating abilities of insects^{3–7}. Leading-edge vortices

form over the wings of both hawkmoths and models during the downstroke. A vortex held above a wing has long been known to be capable of generating lift^{2,13–19}. A transient leading-edge vortex can be formed by sudden changes in flow velocity²⁰ or pitch²¹, and aerodynamic experiments in unsteady flows have shown that at its peak, the vortex can increase the lift coefficient markedly above the steady-state value for a given aerofoil^{2,20,21}. Studies with model insects^{3,6,7} suggest that leading-edge vortices can be quite stable over model insect wings, and may produce a twofold increase in lift. There are at least two other aerodynamic mechanisms involved in insect flight: one associated with wing rotation^{1,3,7} and another associated with wake capture^{3,7}; however, until now the flow features associated with wake capture and rotational mechanisms have not been identified in real insects, even in tethered flight.

There are also qualitative differences between the published structures of the leading-edge vortices formed over two flapper models^{3,5–7}. The flow field around the hawkmoth flapper model is apparently analogous to the leading-edge vortices over delta wings, with vortex stability maintained by the removal of vorticity through a spanwise (base of the wing to the wing-tip) axial flow along the vortex cores^{5,6}. However, chordwise fences on the wings of the *Drosophila* flapper model stopped spanwise flow but not the leading edge vortex, which even increased in size and strength³. Flow visualization experiments with tethered hawkmoths showed leading-edge vortices^{5,6}, but with insufficient resolution to reveal the internal structure. Furthermore the hawkmoths were tethered, and although an insect on a tether may flap its wings, this is not real flight.

To overcome the problems associated with tethered insects, we used a high-resolution, flow visualization system to examine the details of the field of air flow around the wings of free-flying butterflies (*Vanessa atalanta*). To describe the topology of the different three-dimensional separated flows that free-flying butterflies produce, we used critical point concepts^{22–27}.

The butterflies in our experiments used all of the unsteady aerodynamic mechanisms that have been proposed. They switch between mechanisms freely—often using completely different mechanisms on successive wing strokes—and are able to choose different aerodynamic mechanisms to suit different flight behaviours. Visualizations of air flow in free flight show that the fluttering of butterflies is not a random, erratic wandering, but results from the mastery of a wide array of aerodynamic mechanisms available to free-flying insects.

Figure 1a shows the typical flow field over the midline of a butterfly during a downstroke. The flow is clearly not of the type described in previous studies^{3,5,7} and of that shown in Fig. 1b. Instead, an open U-shaped separation exists above the insect with a free-slip critical point (saddle) above the midline. This distinct pattern of leading-edge vortex flow has a nodal point of attachment on the midline. Lines of attachment run from this nodal-point along the undersurface of the wings to the wing-tips; there is a saddle point of separation on the top surface of the body from which lines of separation run to the wing-tips close to, or along, the leading edge. There is a free-slip critical point (saddle) above the midline, and the leading-edge vortices extend out from this free-slip critical point along the wings and on into the wing-tip vortices (with which they are continuous). The leading-edge vortices are not conical or spiral in any of our images, but are of approximately constant diameter, and if there is any spanwise flow it is too weak to be detected. In such images, spanwise flow would appear as a distortion of the curved smoke streams passing round the leading-edge vortex, with streams nearer the centre of the leading-edge vortex being pulled out towards the wing-tips or even turning to flow along the wings towards the wing-tips. This sort of distortion is not visible in any of our flow visualizations of *V. atalanta*. Re-attachment occurs either at a nodal point of attachment on the rear of the body from which lines of attachment run to the wing-tips, or at a free-slip

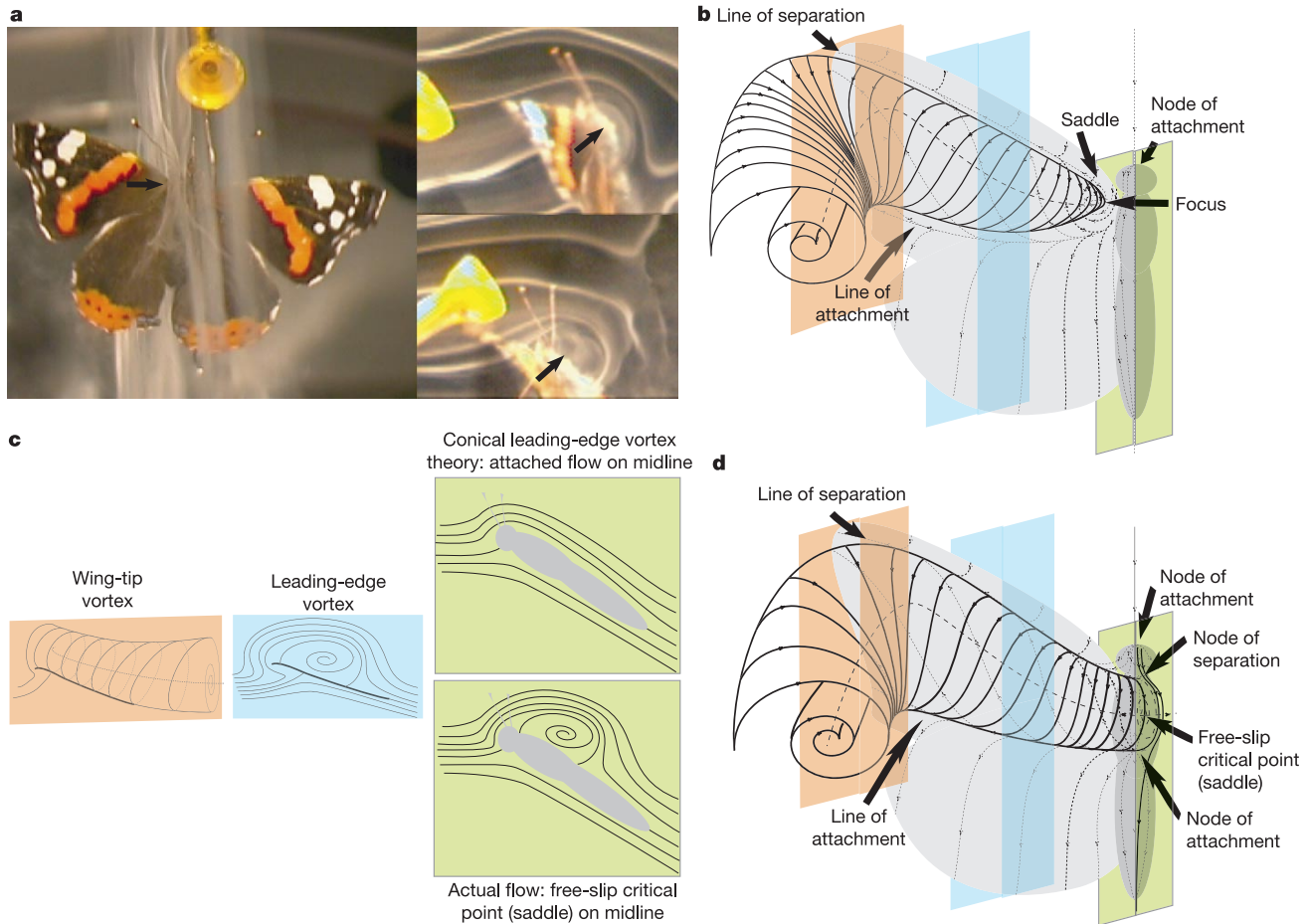


Figure 1 Butterfly leading edge vortex structure. **a**, Smoke-wire visualization of the flow over the midline of *V. atalanta* in free flight. Top and side views of a 397-mg female flying into a 1.5 m s^{-1} wind are shown. The high-speed camera images were taken immediately before and after the high-resolution top view. The images show a free-slip critical point (saddle) above the midline of the animal (arrows). **b**, Critical point reconstruction of the flow described from tethered flight and flapping flight studies^{2,4,5} with conical-spiral leading-edge vortices (Werle-Legendre separations) over each wing and attached flow over the midline. The flow seen in **a** is not consistent with this pattern of flow. **c**, Side views showing the flow that would be visualized by a vertical array of smoke

streams positioned along the midline, half-way out along the wing, and at the wing-tip. The conical-spiral leading-edge vortices would require an attached flow at the midline, which is not seen in *V. atalanta*. **d**, Critical point interpretation of the flow that is seen in flow visualizations of *V. atalanta*. There is a free-slip critical point (saddle) above the midline of the animal, and the cores of the leading-edge vortices run from this saddle point out along the leading edges and into the wing-tip vortices, with which they are continuous. The leading-edge vortices are of constant diameter, and spanwise flow is negligible or absent (see Supplementary Information for video footage and further images showing these details).

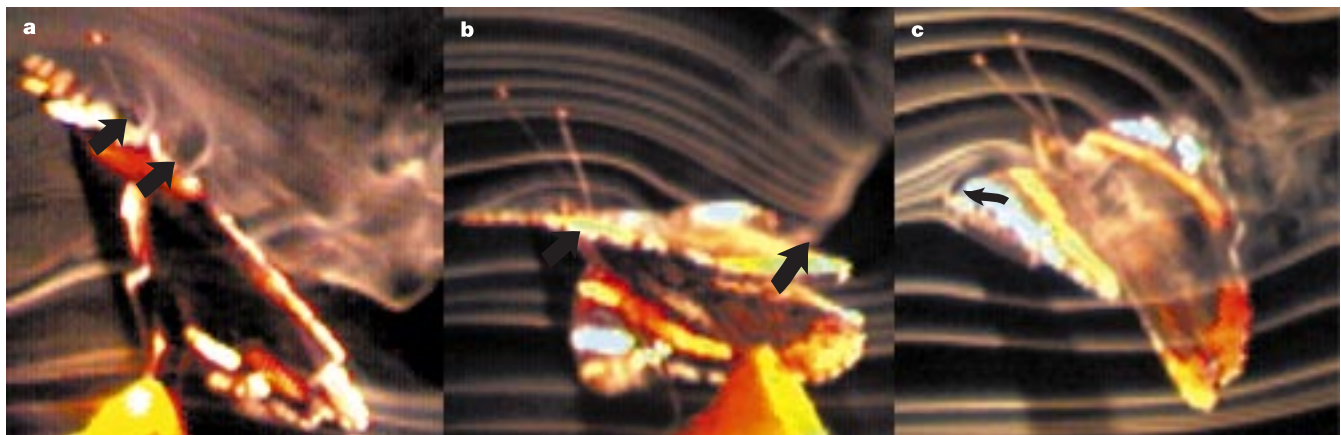


Figure 2 Flow over the wings during the downstroke. **a**, Side view of a double leading-edge vortex (arrows) used to generate extreme accelerations in vertical take off by a 364-mg male flying into a 1.0 m s^{-1} wind. **b**, Attached flow used by a 434-mg male during steady forward flight into a 1.5 m s^{-1} wind. The arrow points to the bright smoke streams flowing straight towards the camera as they wrap around the wing-tip vortex.

c, Leading-edge vortex used during acceleration by the same 434-mg male in the same video sequence. The arrow points to a smoke stream that splits in front of and above the leading edge where it encounters the rapidly rotating air in the leading-edge vortex, which is moving vertically at this point; that is, in a perpendicular direction to the smoke stream arriving at the wing.

critical point (node) behind the body from which bifurcation lines run to the wing-tips. Again saddle points of separation will be present at the wing-tips. Figure 1d is a sketch of the main features and critical points of this type of flow field. Smoke visualization is a qualitative technique which cannot pick up subtle variations in velocity in a flow field, but the huge separation over the midline in Fig. 1a and the clear and unequivocal presence of a free-slip critical point above the midline rules out the conical-spiral leading-edge vortex theory previous work had suggested. It would not be possible for single-winged flappers to accurately reproduce the flow seen in Fig. 1a, so their results^{3,5–7} should be viewed with caution.

The leading-edge vortices are present over the wings from the beginning of the downstroke. *Vanessa atalanta* uses a variant of the Weis-Fogh and Lighthill's clap-and-fling mechanism (the clap and peel) to start the downstroke (usually), and the leading-edge vortices appear fully formed as soon as the wings separate—our visualizations do not reveal any substantial change in the size or form of the leading-edge vortices during the wing-beat (as might be expected in a dynamic stall mechanism). The clap and peel, and the formation of the leading-edge vortices can be seen in the video sequences provided as Supplementary Information.

Our visualizations show *V. atalanta* using the leading-edge vortex mechanism for gentle acceleration or manoeuvres, and for climbing flight; however, our visualizations of steady forward flight or of rapid accelerations show qualitatively different flow patterns. Figure 2 shows three types of flow over the wings of the butterfly during downstrokes.

During wing-beats that result in very large accelerations of the butterfly, our visualizations show *V. atalanta* producing two sub-parallel leading-edge vortices over the top surface of the wing (Fig. 2a). High-speed video footage shows that the two vortices form during two phases of acceleration at the start of the wing-beat—the initial peel at the start of the downstroke and a subsequent acceleration once the wings have separated. They have the same sense of rotation so they are not equivalent to the primary and secondary vortices seen over delta wings²⁰.

In steady forward flight we find no evidence of a leading-edge vortex in *V. atalanta*. The leading-edge vortex is part of a high-lift mechanism, which may also generate large drag. Any fluid captured in the leading-edge vortex loses streamwise velocity as it is accelerated into the flow of the leading-edge vortex^{8,15}. Theoretical^{4,7,8,16} and experimental^{7,13–15} analyses show that the lift to drag ratio is higher under attached flow conditions than when flow is separated to form a leading-edge vortex. High lift is not required for steady forward flight, so the absence of a leading-edge vortex (Fig. 2b) in forward free flight in our experimental butterflies makes sense from an energetics perspective. The existence of a large leading-edge vortex in previous work with hawkmoths and particularly its increase in size in fast forward flight²⁸ is puzzling, and may be an artefact of tethering. Flow visualization with free-flying hawkmoths is required.

During the final stages of the downstroke in *V. atalanta* there is often (but not always) a phase where the wing stops translating, and instead rotates in a nose-up sense. The aerodynamics of rotational mechanisms are well known for cylinders rotating in a steady flow^{1,2}. Wing rotation can also produce circulation, and this rotational circulation adds to the normal translational circulation^{1–3}. This is demonstrated most clearly in one of the video sequences (see Supplementary Information) at the end of the second wing-beat where the size of the wing-tip vortex increases substantially as the wings rotate rapidly at the end of the downstroke.

At the ends of the wing strokes when the wing stops, circulation is shed into the wake in the form of discrete energetic stopping vortices. A wing passing through a stopping vortex could extract energy from it if correctly orientated^{3,7}. Optimally, to save energy, wake capture would increase the size of the stopping vortex (increasing the mass flow in the wake) while reducing the circulation of the stopping vortex (reducing the velocity in the wake). The same momentum (mv) could be shed into the wake at a lower cost in terms of kinetic energy ($1/2 mv^2$). Alternatively, wake capture could involve the extraction of circulation from the stopping vortex to form the bound vortex on the wing for the next stroke, or to

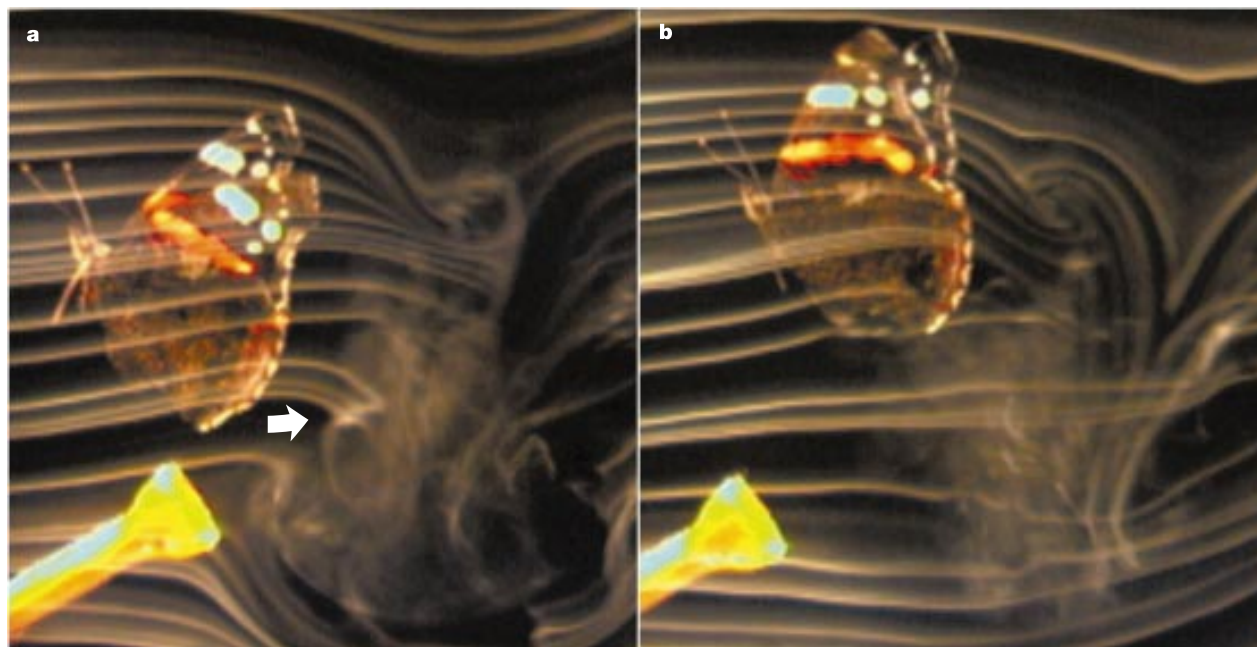


Figure 3 Wake capture. The images show successive wing-beats by a 434-mg male flying into a 1.5 m s^{-1} wind. **a**, The wing-beat has produced a wake structure with a clearly defined stopping vortex marked by the curved smoke streams at the arrow. No wake capture has occurred. **b**, Instead of a small, high-velocity stopping vortex marked by

curved smoke streams behind and below the insect, there is a large area of disorganized smoke at the lower right of the image. The wing has moved through the stopping vortex on the subsequent upstroke, disrupting the structure of the stopping vortex.

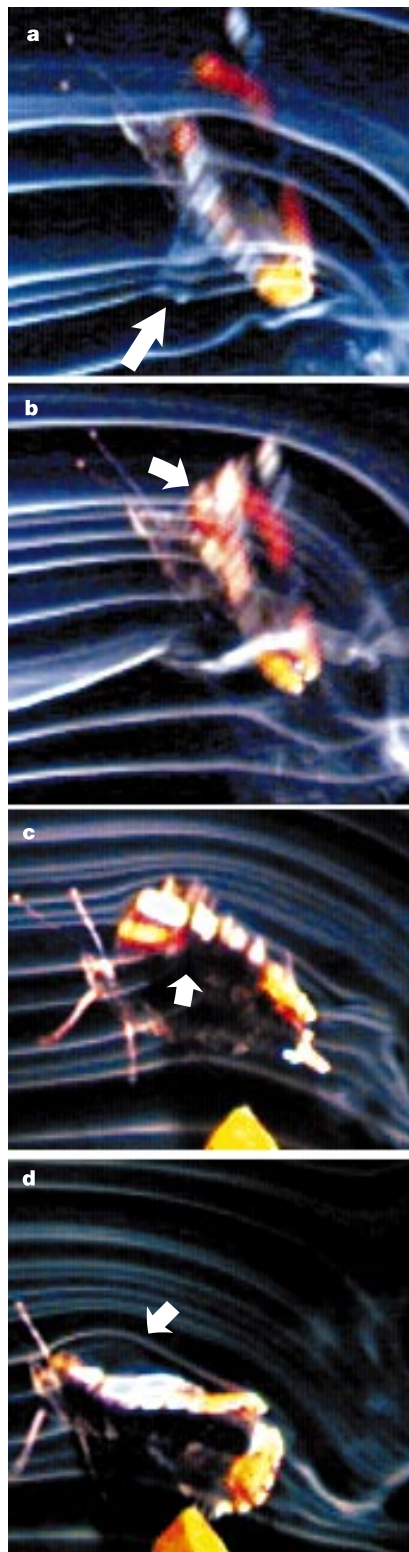


Figure 4 Upstrokes with positive, neutral and negative (downwards) loadings produced during a single flight by a 434-mg male flying into a 1.5 m s^{-1} wind. **a**, Disjointed smoke streams indicative of a downward-loaded upstroke, and the wing-tip vortex is clearly visible (arrow). The upstroke in this case would produce forwards and downwards forces. **b**, The smoke streams are undisturbed by the passage of the wing (arrow), indicating that the wing is unloaded. **c**, A positively loaded upstroke, and the curvature of the smoke streams close to the stagnation point (arrow), indicate positive circulation around the wing. **d**, Positive circulation and a leading-edge vortex (arrow) on the upstroke. This upstroke would be generating forces directed upwards and backwards.

accelerate the wing by the interaction of the near-stationary wing with the accelerated flow around it (added mass force).

Figure 3 shows the wake at the late upstroke stage in two wing-beats—one with wake capture (Fig. 3b) and one without (Fig. 3a). The effect of wake capture is clear in the difference between these frames. When the upstroke is configured so that the wing interacts with the stopping vortex, the structure of the stopping vortex is disrupted, increasing the mass of fluid in motion but decreasing the intensity of its rotation. The flow visualizations in Fig. 3 are entirely consistent with the use of wake capture as a way of increasing the efficiency of the flight mechanism by increasing the mass at the same time as reducing the velocity in the stopping vortex.

Free-flying butterflies use a wide array of different aerodynamic mechanisms in flight. In successive wing-beats, *V. atalanta* can choose to use, or not to use, wake capture, leading-edge vortices, a clap and peel mechanism, and whether to use positive, negative or zero loading on the upstroke (Fig. 4; see also Supplementary Information).

In birds and bats the loading on the upstroke changes from negative or zero at low speeds to positive at higher speeds in a manner analogous to the gait changes of terrestrial animals^{29,30}. *Vanessa atalanta*, in contrast, uses radically different aerodynamic mechanisms on successive wing-beats, rather than successive gaits at different speeds.

Our results show a clear correlation between the aerodynamic mechanism that *V. atalanta* selects and the acceleration produced by that wing-beat. *Vanessa atalanta* selects leading-edge vortex mechanisms, clap and peel, and rotational mechanisms when it requires very high lift forces, and avoids them in steady forward flight.

The micro-air-vehicle community may find it daunting that the first flow visualizations of free-flying insects have revealed such a wide range of aerodynamic mechanisms, and that the insect switches between them on successive wing strokes with apparent ease. The situation may not be as complex as it first seems, because inspection of the flow visualization videos suggests that *V. atalanta* switches between different aerodynamic mechanisms through rather simple changes in wing-beat kinematics. For example, the difference between wing-beats with and without wake capture is that in wake-capture events there is less anterior motion of the wing at the end of the downstroke. Relatively subtle changes in kinematics can change radically the aerodynamic mechanism that the insect uses. The rapid shift between aerodynamic mechanisms and consequent large changes in the forces produced per wing stroke may be why butterflies flutter. □

Methods

Equipment and experimental design

Butterflies were captured at the Oxford University Zoology Department and trained to feed from an artificial feeder (golf tee). The feeder was then positioned in a low-turbulence, low-noise wind tunnel operating at a range of speeds from 0.5 to 2.5 m s^{-1} . Butterflies were filmed after feeding and shivering to warm up; hence, all video sequences used in the analyses were of natural, self-motivated free flight. Smoke streams were generated by the smoke-wire technique (Horners steam engine oil on electrically heated 0.1-mm nichrome wire). An array of direct current spotlights giving a total of 650 W provided even overhead illumination. Butterflies were filmed using two Canon XL1 camcorders, four Canon MV30DV camcorders and high-speed video (NAC500, 250 frames per second). Multiple cameras were used to provide stereo and orthogonal views to permit reconstruction of the highly three-dimensional, unsteady flows produced by the insects. Images presented here are from Vanessa11, Vanessa5 and Vanessa4 sequences. Quicktime versions of the original video and a set of supplementary figures are available as Supplementary Information. The images presented here are taken from the original, undistorted 640×480 pixel digital video images. No image manipulation has been performed other than the addition of arrows or labels to the figures.

Critical point theory

In unsteady separated flows, streamlines converge on or diverge from particular points, and the streamline vector is undefined at those singular critical points^{24–27}. For example, where the flow attaches at the front of a sphere there is a single attachment point and streamlines on the surface radiate out from this critical point, which is called a node^{23,27}. The direction of the flow at the nodal point of attachment is undefined. The other

common types of critical point are foci^{23,27} (such as the point on the surface where the centre of a vortex touches down) and saddles^{23,27} (which have two streamlines entering and two leaving the critical point from opposite directions). Saddle points are thought to be characteristic of separated flows^{21–24}. The distribution of critical points, and of the streamlines joining them, defines the topology of the flow, and obeys topological rules^{23,24}. For example, the number of nodes plus foci must equal the number of saddles plus 2 in the skin friction lines on a body surface, just as faces plus corners equals edges plus 2 for a solid body such as a cube.

Received 26 March; accepted 7 October 2002; doi:10.1038/nature01223.

1. Sane, S. P. & Dickinson, M. H. The aerodynamic effects of wing rotation and a revised quasi-steady model of flapping flight. *J. Exp. Biol.* **205**, 1087–1096 (2002).
2. Zbikowski, R. On aerodynamic modelling of an insect-like flapping wing in hover for micro air vehicles. *Phil. Trans. R. Soc. Lond. A* **360**, 273–290 (2002).
3. Dickinson, M. H., Lehmann, F.-O. & Sane, S. P. Wing rotation and the aerodynamic basis of insect flight. *Science* **284**, 1954–1960 (1999).
4. Maxworthy, T. The fluid-dynamics of insect flight. *Ann. Rev. Fluid Mech.* **13**, 329–350 (1981).
5. Ellington, C. P., van den Berg, C., Willmott, A. P. & Thomas, A. L. R. Leading-edge vortices in insect flight. *Nature* **384**, 626–630 (1996).
6. Van den Berg, C. & Ellington, C. P. The vortex wake of a hovering model hawkmoth. *Phil. Trans. R. Soc. Lond. B* **352**, 317–328 (1997).
7. Birch, J. M. & Dickinson, M. H. Spanwise flow and the attachment of the leading-edge vortex on insect wings. *Nature* **412**, 729–733 (2001).
8. Saffman, P. G. & Sheffield, J. S. Flow over a wing with an attached free vortex. *Studies Appl. Math.* **57**, 107–117 (1977).
9. Weis-Fogh, T. Quick estimates of flight fitness in hovering animals, including novel mechanisms for lift production. *J. Exp. Biol.* **59**, 169–230 (1973).
10. Lighthill, J. On the Weis-Fogh mechanism of lift generation. *J. Fluid Mech.* **60**, 1–17 (1973).
11. Maxworthy, T. Experiments on the Weis-Fogh mechanism of lift generation by insects in hovering flight. Part 1. Dynamics of the 'filing'. *J. Fluid Mech.* **93**, 47–63 (1979).
12. Ellington, C. P. *Biological Fluid Dynamics Symp. Soc. Exp. Biol.* (eds Ellington, C. P. & Pedley, T. J.) Vol. 49, 109–129 (Company of Biologists, Cambridge, 1995).
13. Rossow, V. J. Lift enhancement by an externally trapped vortex. *J. Aircraft* **15**, 618–625 (1978).
14. Rossow, V. J. Two-fence concept for efficient trapping of vortices on airfoils. *J. Aircraft* **29**, 847–855 (1992).
15. Rossow, V. J. Aerodynamics of airfoils with vortex trapped by two spanwise fences. *J. Aircraft* **31**, 146–153 (1994).
16. Huang, M.-K. & Chow, C.-Y. Trapping of a free vortex by Joukowski airfoils. *Am. Inst. Aeronaut. Astronaut. J.* **20**, 292–298 (1982).
17. Mourtos, N. J. & Brooks, M. Flow past a flat plate with a vortex/sink combination. *J. Appl. Mech.* **63**, 543–550 (1996).
18. Riddle, T. W., Wadcock, A. J., Tso, J. & Cummings, R. M. An experimental analysis of vortex trapping techniques. *J. Fluids Eng.* **121**, 555–559 (1999).
19. Gursul, I. & Ho, C.-M. High aerodynamic loads on an airfoil submerged in an unsteady stream. *Am. Inst. Aeronaut. Astronaut. J.* **30**, 1117–1119 (1992).
20. Gad-el-Hak, M. & Ho, C.-M. Unsteady vortical lift around three-dimensional lifting surfaces. *Am. Inst. Aeronaut. Astronaut. J.* **24**, 713–721 (1986).
21. Déleury, J. M., Legendre, R. & Werlé, H. Toward the elucidation of three-dimensional separation. *Ann. Rev. Fluid Mech.* **33**, 129–154 (2001).
22. Perry, A. E. & Chong, M. S. A description of eddy motions and flow patterns using critical-point concepts. *Ann. Rev. Fluid Mech.* **19**, 125–155 (1987).
23. Tobak, M. & Peake, D. J. Topology of three-dimensional separated flows. *Ann. Rev. Fluid Mech.* **14**, 61–85 (1982).
24. Lighthill, M. J. *Laminar Boundary Layer Theory Section II 2.6* (ed. Rosenhead, L.) 72–82 (Oxford Univ. Press, New York, 1963).
25. Legendre, R. Separation de l'écoulement laminaire tridimensionnel. *Rech. Aeronaut.* **54**, 3–8 (1956).
26. Poincaré, H. Les points singuliers des équations différentielles. *C.R. Acad. Sci. Paris* **94**, 416–418 (1882).
27. Hornung, H. & Perry, A. E. Some aspects of three-dimensional separation. I: Stream surface bifurcations. *Z. Flugwiss. Weltraumforsch.* **8**, 77–87 (1984).
28. Willmott, A. P., Ellington, C. P. & Thomas, A. L. R. Flow visualisation and unsteady aerodynamics in the flight of the hawkmoth *Manduca sexta*. *Phil. Trans. R. Soc. Lond. B* **352**, 303–316 (1997).
29. Rayner, J. M. V., Jones, G. & Thomas, A. L. R. Vortex flow visualizations reveal change in upstroke function with flight speed in bats. *Nature* **321**, 162–164 (1986).
30. Spedding, G. R. *Advances in Comparative Environmental Physiology 11. Mechanics of Animal Locomotion* (ed. Alexander, R. M.) (Springer, Berlin, 1993).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank the Engineering and Physical Sciences Research Council instrument pool for use of their NAC500 high-speed video camera. R.B.S. was supported by a Biotechnology and Biological Sciences Research Council grant to A.L.R.T. A.L.R.T. was supported by a Royal Society University Research Fellowship.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to A.L.R.T. (e-mail: adrian.thomas@zoo.ox.ac.uk).

Sex releases the speed limit on evolution

Nick Colegrave

Institute of Cell, Animal and Population Biology, University of Edinburgh, Edinburgh EH9 3JT, UK

Explaining the evolutionary maintenance of sex remains a key problem in evolutionary biology^{1–3}. One potential benefit of sex is that it may allow a more rapid adaptive response when environmental conditions change, by increasing the efficiency with which selection can fix beneficial mutations^{4–7}. Here I show that sex can increase the rate of adaptation in the facultatively sexual single-celled chlorophyte *Chlamydomonas reinhardtii*, but that the benefits of sex depend crucially on the size of the population that is adapting: sex has a marked effect in large populations but little effect in small populations. Several mechanisms have been proposed to explain the benefits of sex in a novel environment, including stochastic effects in small populations, clonal interference and epistasis between beneficial alleles. These results indicate that clonal interference is important in this system.

As pointed out by Fisher⁴, for sex to increase the rate of adaptation requires that the supply of beneficial mutations be abundant. If mutations arise only rarely, each is fixed before the next arises, and populations spend most of their time waiting for new mutations. Sex has little benefit under such conditions. If beneficial mutations arise more commonly, several different mutations are spreading through the population at the same time. Under such conditions, theoretical models show that clonal interference in asexual populations might place an important speed limit on adaptation, and sex should provide a significant benefit⁸. The robustness of this prediction depends to some extent on the unknown distribution of beneficial mutations, because current models of clonal interference ignore the possibility that multiple beneficial alleles arise in the same lineage^{8,9}. If only a very small number of beneficial mutations (for example, two) are possible in the new environment, then the probability of both of these arising in a single asexual individual becomes non-trivial in very large populations. However, the probability that all mutations might arise simultaneously declines steeply and rapidly with the number of possible beneficial mutations. In this situation it seems likely that the genotypes carrying different subsets of beneficial alleles interfere in the same way as for a single mutation⁹. Furthermore, there is experimental evidence that clonal interference does occur in asexual populations of bacteria and viruses^{9,10}.

Because large populations have a larger number of mutations per generation than small populations, we might predict that the benefits of sex are greater in large populations. However, the effect of population size also depends on whether benefits of sex are based on stochastic associations between the beneficial mutations, which

Table 1 **Effective population size for the different experimental treatments**

Bottleneck size (N_0)	Approximate number of generations between transfers (g)	N_e
10^3	15	15,000
10^4	12	120,000
10^5	8.5	850,000
5×10^5	6.5	3,250,000
10^6	5.5	5,500,000

In common with other such microbial studies, I estimate the effective population size for beneficial substitutions (N_e) from an estimate of the number of divisions that occur in each tube between transfers (g), and the original inoculum size (N_0). $N_e = N_0 g$ (ref. 15).

are important in small populations¹¹, on clonal interference (which despite operating most strongly in large populations is based on stochastic processes) or on associations due to deterministic forces such as epistasis that occur at all population sizes¹¹. Because the net effect of sex is affected by all of these factors, the actual way in which population size affects the benefits of sex is unclear and requires experimental investigation.

I allowed 20 asexual populations of *C. reinhardtii* to evolve for 250 generations in a simple laboratory environment containing a novel growth medium. Population size was manipulated by regularly bottlenecking populations to different degrees (to between 10^3 and 10^6 cells; see Table 1 for bottleneck and effective population sizes). I then examined the level of adaptation of these evolved populations by comparing their growth rate with the growth rate of their ancestor under the conditions of selection.

To examine the interaction between sex and population size, samples were taken from each of the asexual populations in three of the bottleneck treatments (10^3 , 10^5 and 10^6) after 150 generations. Each population was passed through eight rounds of sexual reproduction and then allowed to evolve for a further 50 generations. During this subsequent evolution, all populations, both sexual and asexual, were bottlenecked to the same size (10^4 cells) at weekly

intervals, to avoid confounding effects of sex with subsequent differences in population size. Unmated samples were also run under the same regime. The relative fitness of the sexual populations was then determined by comparing the growth rate of each sexual population with that of its asexual control.

In the asexual lines, bottleneck size affected the rate of adaptation, with a linear relationship between $\log_{10}(\text{adaptation})$ and $\log_{10}(\text{effective population size})$ ($F_{1,15} = 8.54$, $P = 0.010$; Fig. 1a). Moreover, there was no evidence for any difference in the effect of population size across lines ($F_{3,15} = 2.00$, $P = 0.157$), or of any overall difference between the rate at which the different lines adapted ($F_{3,12} = 0.65$, $P = 0.615$). The slope of the relationship was less than 1 (0.19 ± 0.06 (mean $\pm$ s.e.), $t = 13.5$, d.f. = 15, $P < 0.001$), implying a diminishing-returns relationship on the untransformed scale, consistent with clonal interference (Fig. 1b).

Sex produced a general increase in the rate of adaptation, with the overall relative fitness of the sexual lines being greater than 1 (Fig. 2; pooled relative fitness = 1.10 ± 0.02 (mean $\pm$ s.e.), $t = 5$, d.f. = 6, $P < 0.01$). More importantly, the magnitude of the benefit of sex depends on the size of the population before the induction of sex. The largest population showed a substantial benefit of sex, whereas the smallest populations differed little from the asexual controls ($F_{2,6} = 5.92$, $P = 0.038$). Thus, in large populations sex seems to be able to release the speed limit on adaptation set by clonal interference.

Although the effect of population size on adaptation has been previously investigated in asexual populations of bacteria and viruses^{9,10}, and I have shown previously that sex can increase the rate of adaptation in *C. reinhardtii*¹², this is to my knowledge the first study that has examined experimentally the interaction between sex and population size on the rate of adaptation. The results of this experiment provide striking support for the conjecture that sex has the greatest effect in large populations where the supply of beneficial mutations is plentiful^{2,4}. In a small population, the process of adaptation might be limited by the supply of novel mutations, and sex is of little benefit, whereas in large populations the limit might be imposed by the efficiency of selection, which can be markedly increased by sex.

The greater benefit of sex in the large population might at first seem counter to many people's intuition, and much theory, which shows that advantages of sex often depend on stochastic effects in small populations¹⁻³. However, clonal interference, despite being based ultimately on stochastic forces (that is, the absence of some genotypes in finite populations), only becomes an important limit to the rate of adaptation in larger populations. The results of this study show that clonal interference can be important in providing a benefit to sex, in this system at least, and that the increased efficiency of selection might provide a benefit of sexual reproduction even in organisms with large population sizes. □

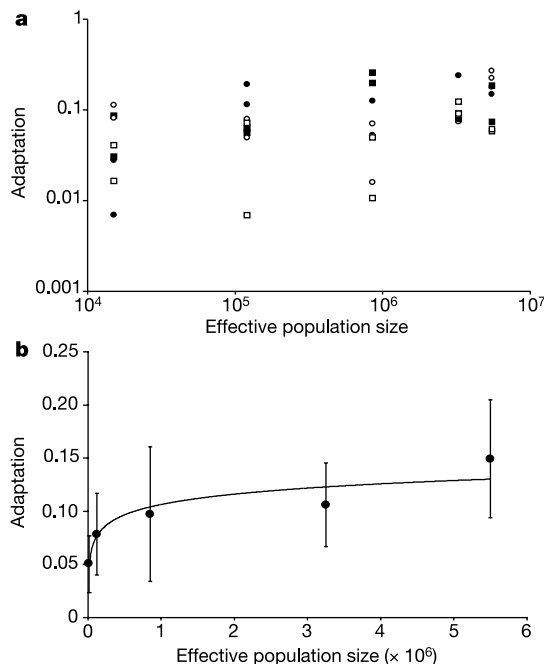


Figure 1 Effect of bottleneck size on adaptation in asexual populations. **a**, Individual points represent individual assays, and the different symbols indicate the lines from the initial four base populations. The relationship was examined statistically by fitting general linear models (GLM) to the data after log-transforming both variables. This has several advantages over using nonlinear regression on the untransformed data: the GLM framework means that line effects and line-by-population-size interactions can be fitted to the data as well as the population size term; and working with logarithms of population sizes reduces problems of leverage by the extreme population sizes. Initially I fitted both quadratic and linear population size terms, but the quadratic term did not significantly improve the fit, and so was dropped from the model ($F_{1,11} = 0.16$, $P = 0.69$). **b**, Overall pooled means and standard errors of the lines on an untransformed scale along with the fitted relationship from the GLM. On the untransformed scale the linear relationship takes the form $\text{adaptation} = b(N_e)^c$, a relationship that is potentially able to describe a variety of relationships from diminishing returns to linear to accelerating. The actual fitted relationship is $\text{adaptation} = 0.00575 (\text{effective population size})^{0.19}$, which is clearly of the diminishing-returns form. I chose a relationship that passes through the origin because a population of size zero is not expected to adapt.

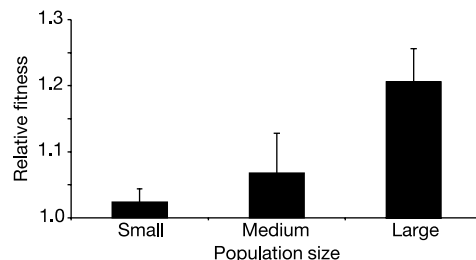


Figure 2 Interaction between sex and bottleneck size. Bars show the mean relative fitness ($\pm$ s.e.) of the sexual lines compared with their asexual controls. Sex has a greater positive effect in the populations that were bottlenecked the least.

Methods

Selection lines were serially propagated in glass tubes containing 10 ml of Bold's medium supplemented with 500 mg l⁻¹ sodium bicarbonate, a substance that reduces the growth rate of *C. reinhardtii*. Tubes were maintained under constant lighting and shaken continuously. Temperature was not controlled. The experiment was begun with four base populations each derived (about 30 generations previously) from a single different *mt*⁺ spore drawn from a heterogeneous population. Each base population will have contained minimal genetic diversity. A sample of each base population was used in each of the five bottleneck treatments, giving a total of 20 experimental lines.

Bottlenecking was achieved by transferring a known number of cells to the new tube of medium after 7 days of growth. Cell density was estimated before transfer by measuring the absorbance of the culture at 665 nm (a measurement that is a good predictor of cell density over the range of densities used in this experiment (unpublished observations)). Using bottlenecks to vary population size is the method that has typically been used in studies of microbial population size. It has the disadvantage that it affects the selective environment as well as the population size. Such differences will be unimportant if the major response to selection is through an increase in exponential growth rate, as in other microbial systems¹³. I minimized any possible difference by choosing bottleneck sizes and volumes of medium where the populations were still growing at the time of transfer, and had not reached carrying capacity. Furthermore, there was a significant positive correlation between the fitness of the evolved lines from the different bottleneck treatments when assayed under the lowest and highest density conditions ($r = 0.89$, $n = 20$, $P < 0.001$) and also between the early growth rate of lines and their population size on day 7 ($r = 0.96$, $n = 20$, $P < 0.0001$).

To examine the effects of sex, samples of the populations of 10⁶ cells were placed into nitrogen-free medium for 12 h to induce gamete production. Because all lines used in the experiment were of a single mating type it was then necessary to add cells of another clone of the opposite mating type for the first round of sex. All populations were mated with the same clone. Cells were placed into the light for 12 h to allow mating to occur, after which point mats of zygotes were clearly visible. Zygotes were matured in the dark for 5 days on solid medium and then germinated in the light for 5 days. The resulting populations, now containing a mixture of mating types, were then mated a further seven times in the same way (without the addition of any more cells). Such a procedure does involve the introgression of DNA from an unevolved line, which might affect adaptation, but it should do so to the same extent in all populations. Multiple matings were used because in the first mating recombination is only possible between mutations that have arisen within a selection line and mutations from the new genotype. Only by allowing multiple sexual cycles is it possible to allow recombination between the mutations that arose during selection. Asexual controls underwent the same sampling and bottlenecking regime as the sexual lines and were treated in the same way, with the exception that cells of the opposite mating type were never added, and cells were maintained in dim light rather than darkness for the development stage. Nitrogen starvation has been shown to produce a small increase in the mutation rate in *C. reinhardtii*¹⁴, and although this led to no measurable effect on rates of adaptation¹² in an experiment similar to this, starving both the sexual and asexual populations to the same extent ensured that any such effect did not confound the effects of sex.

The level of adaptation of a line was estimated by using a paired design. Pairs of tubes containing the experimental growth medium were inoculated with a known number (corresponding to the appropriate bottleneck size) of either the evolved line or its ancestor. Thus, lines were assayed under the same conditions as those in which they had evolved. Two replicate pairs of tubes were set up for each evolved line. Pairs of tubes were allocated randomly to a position on the growth shelves, but the tubes in a pair were always placed next to one another. The fitness of a line was defined as the number of cells after 7 days of growth (again determined by absorbance at 665 nm), and the amount of adaptation of a line as (evolved fitness – ancestors' fitness)/ancestors' fitness.

Assaying populations under their specific selection conditions meant that lines in different treatments were assayed under different conditions. However, similar results were obtained when populations were assayed under the same conditions (all inoculated with 10⁴ cells; data not shown). To examine the effects of sex on rate of adaptation, the same paired assay was used, except that one tube of each pair contained the sexual line to be investigated, and the other its asexual control. Three replicate pairs were set up for each sexual line, but three of the replicates were lost owing to contamination.

All statistical analyses in this paper were performed on the means of the replicate measures for each selection line, to avoid problems of non-independence.

Received 9 June; accepted 24 September 2002; doi:10.1038/nature01191.

- Maynard Smith, J. *The Evolution of Sex* (Cambridge Univ. Press, Cambridge, 1978).
- Bell, G. *The Masterpiece of Nature* (Croom Helm, London, 1982).
- West, S. A., Lively, C. M. & Read, A. F. A pluralist approach to sex and recombination. *J. Evol. Biol.* **12**, 1003–1012 (1999).
- Fisher, R. A. *The Genetical Theory of Natural Selection* (Clarendon, Oxford, 1958).
- Muller, H. J. Some genetic aspects of sex. *Am. Nat.* **66**, 118–138 (1932).
- Otto, S. P. & Barton, N. H. The evolution of recombination: removing the limits to natural selection. *Genetics* **147**, 879–906 (1997).
- Burt, A. Perspective: Sex, recombination, and the efficacy of selection—was Weismann right? *Evolution* **54**, 337–351 (2000).
- Gerrish, P. J. & Lenski, R. E. The fate of competing mutations in an asexual population. *Genetica* **102/103**, 127–144 (1998).
- de Visser, J. A. G. M., Zeyl, C. W., Gerrish, P. J., Blanchard, J. L. & Lenski, R. E. Diminishing returns from mutation supply rates in asexual populations. *Science* **283**, 404–406 (1999).
- Miralles, R., Gerrish, P. J., Moya, A. & Elena, S. F. Clonal interference and the evolution of RNA viruses. *Science* **285**, 1745–1747 (1999).

- Otto, S. P. & Barton, N. H. Selection for recombination in small populations. *Evolution* **55**, 1921–1931 (2001).
- Colegrave, N., Kaltz, O. & Bell, G. The ecology and genetics of fitness in *Chlamydomonas*. VIII. The dynamics of adaptation to novel environments after a single episode of sex. *Evolution* **56**, 14–21 (2002).
- Vasi, E., Travisano, M. & Lenski, R. E. Long-term experimental evolution in *Escherichia coli*. 2. Changes in life-history traits during adaptation to a seasonal environment. *Am. Nat.* **144**, 432–456 (1994).
- Goho, S. & Bell, G. Mild environmental stress elicits mutations affecting fitness in *Chlamydomonas*. *Proc. R. Soc. Lond. B* **267**, 123–129 (2000).
- Lenski, R. E., Rose, M. R., Simpson, S. C. & Tadler, S. C. Long-term experimental evolution in *Escherichia coli*. 1. Adaptation and divergence during 2,000 generations. *Am. Nat.* **138**, 1315–1341 (1991).

Acknowledgements I thank N. Barton, G. Bell, T. Johnson and S. Nee for discussion of this experiment and for comments on the manuscript; A. Poon for suggestions for improvement to earlier versions; S. Otto and A. Read for advice at early stages; and D. Haydon, L. Kruuk and M. Spencer for lengthy discussions on curve fitting. This work was supported by a NERC postdoctoral fellowship.

Competing interests statement The author declares that he has no competing financial interests.

Correspondence and requests for materials should be addressed to N.C. (e-mail: n.colegrave@ed.ac.uk).

Large clusters of co-expressed genes in the *Drosophila* genome

Alexander M. Boutanaev*, Alla I. Kalmykova†, Yuri Y. Shevelyov† & Dmitry I. Nurminsky*

* Department of Anatomy & Cell Biology, Tufts University School of Medicine, Boston, Massachusetts 02111, USA

† Department of Molecular Genetics of Animals, Institute of Molecular Genetics, Russian Academy of Sciences, Moscow 123182, Russia

Clustering of co-expressed, non-homologous genes on chromosomes implies their co-regulation. In lower eukaryotes, co-expressed genes are often found in pairs^{1,2}. Clustering of genes that share aspects of transcriptional regulation has also been reported in higher eukaryotes^{3,4}. To advance our understanding of the mode of coordinated gene regulation in multicellular organisms, we performed a genome-wide analysis of the chromosomal distribution of co-expressed genes in *Drosophila*. We identified a total of 1,661 testes-specific genes, one-third of which are clustered on chromosomes. The number of clusters of three or more genes is much higher than expected by chance. We observed a similar trend for genes upregulated in the embryo and in the adult head, although the expression pattern of individual genes cannot be predicted on the basis of chromosomal position alone. Our data suggest that the prevalent mechanism of transcriptional co-regulation in higher eukaryotes operates with extensive chromatin domains that comprise multiple genes.

We initially observed that five non-homologous, testes-specific genes (*Crtp*, *Yu*, *CK2βtes*, *Pros28.1B* and *CG13581*) form a cluster on chromosome 2 of *Drosophila melanogaster*, implying the presence of large clusters of co-expressed genes in *Drosophila* (A.I.K., D.I.N. & Y.Y.S., unpublished data). To assay the entire *D. melanogaster* genome for the presence of comparable clusters, we used a publicly available EST (expressed sequence tag) database as a source of gene expression data. The content of the EST pool for a given tissue type reflects the composition of original messenger RNA samples used for creation of the complementary DNA library. We generated the software that, through the BLAST homology search, assigns each EST from the database to a gene in the annotated *D. melanogaster* genome, and outputs a table with the numbers of ESTs found for

each gene (for details of the procedure, see Supplementary Information A). As the genes are listed in chromosomal order, the tables (Supplementary Information B) describe an EST representation profile of the genome. A graphical view of an EST representation profile for adult testes is shown for six cytological sections of chromosome 2 (Fig. 1a). Profiles such as these are a model of the transcriptional profile of the genome, providing that the number of ESTs analysed is sufficient. Mapping of 13,979 unique testes-derived ESTs identified about one-third (4,271) of *Drosophila* genes as testes-expressed.

To identify testes-specific genes, we needed to eliminate the genes expressed in other tissue types. For this, we created EST representation profiles for embryos, larvae/pupae, adult ovaries and adult heads, and subtracted all of them from the testes EST profile (Fig. 1b). Out of the 4,271 testes-expressed genes, about two-thirds were also transcribed in other tissues and were eliminated by subtraction, resulting in identification of 1,661 testes-specific genes. To evaluate the reliability of the EST profiling, we analysed the detection rate for 30 testes-specific genes known from the literature and from our own unpublished data. A total of 27 of these were found among 4,271 testes-expressed genes (90% detection rate), and 25 (83%) were included in the pool of 1,661 testes-specific genes after subtraction. Therefore, elimination of a large number of ubiquitously expressed genes by subtraction is very selective, and does not significantly affect the detection of tissue-specific genes.

Next, we analysed tissue specificity of a further 48 genes in three arbitrarily chosen chromosomal regions where clusters of testes-specific genes were detected by EST profiling. The profiles of tissue specificity obtained using real-time polymerase chain reaction with reverse transcription (RT-PCR) (Fig. 2a–c) and by EST profiling (Fig. 2d–f) were similar across the regions, with a correlation coefficient between EST-generated and real-time RT-PCR data of 0.9. Out of 33 testes-specific genes identified by RT-PCR (that is, the genes upregulated more than sixfold in testes as compared with embryos) 24 (73%) were detected by EST profiling. Conversely, out of 26 genes identified as testes-specific by EST profiling, 24 (92%)

were confirmed by the RT-PCR analysis. We then analysed 25 genes in two other genomic regions with a predicted low density of testes-specific genes. Out of two testes-specific genes identified by EST profiling, both were confirmed by RT-PCR and two others were revealed (see Supplementary Information C for RT-PCR data). Therefore, detection of testes-specific genes by EST profiling is very stringent.

Clustering of testes-specific genes was discernible along the chromosomes (Figs 1 and 2). To investigate whether the observed distribution of genes differed from a random distribution, we generated a model of stochastic distribution using a random number generator. EST profiling identified 1,661 testes-specific genes, as described above. The stochastic distribution was generated by producing 1,661 random, non-repetitive numbers in the range of 1–13,290 (the total number of genes in the *Drosophila* genome). With the assumption that the row of numbers from 1 to 13,290 comprises the order of genes in the genome of *D. melanogaster*, each iteration therefore assigns random genomic positions to the 1,661 testes-specific genes. The proportion of genes found in clusters and the size distribution of clusters were calculated, and the values were averaged for 50 reiterations.

Differences between the stochastic distribution (STD) and the distribution revealed by EST profiling (EPD) were observed both in the number of genes included in clusters and in cluster size. Approximately 35% (EPD) of testes-specific genes are clustered; this figure significantly differs from the roughly 24% of clustered genes expected for the STD ($P \ll 0.001$). The EPD size distribution of the gene clusters also differs from the STD prediction ($P \ll 0.001$; Fig. 3), especially for clusters of three or more genes. Although the number of two-gene clusters in the EPD prediction was only 6% higher than expected by chance, there were 2.5 times more three-gene clusters and five times more four-gene clusters in the EPD as compared with the STD prediction. Four five-gene and four six-gene clusters were observed with the EPD; however, clusters of this size were not predicted to form by chance.

The clustering of genes to a greater extent than is predicted by chance may reflect underlying mechanisms of transcriptional co-regulation. A total of 187 clustered testes-specific genes were found by EPD in excess of that expected for STD, of which 90% are

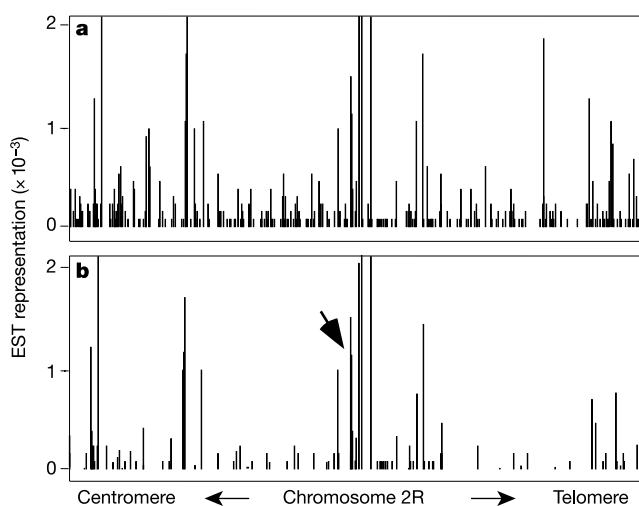


Figure 1 EST profiling reveals clusters of testes-specific genes. For details of the EST mapping procedure and for raw data, see Supplementary Information A and B. **a**, Testes EST representation profile for the cytological sections 50–55 of chromosome 2. The vertical bars represent the genes in their chromosomal order, with the height of the bar representing the proportion of ESTs found for the gene in the testes EST database. **b**, Testes specificity profile obtained by the sequential subtracting of the embryo, larvae/pupae, head and ovary EST profiles from the testes EST profile; only the positive part of the histogram is shown. The arrow indicates the region 53B.

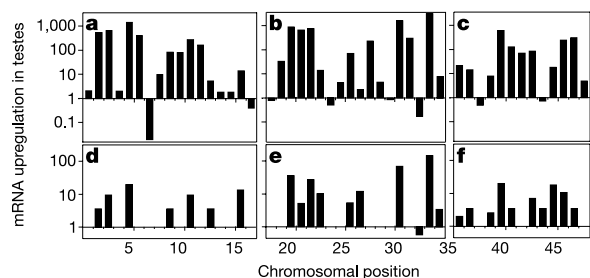


Figure 2 EST profiling and real-time RT-PCR produce similar tissue specificity profiles. Testes specificity of genes in three chromosomal regions, 19E (**a**, **d**), 53B (**b**, **e**) and 60D (**c**, **f**), was determined by real-time RT-PCR (**a**–**c**) and by EST profiling (**d**–**f**). The bars represent the upregulation of the mRNAs in adult testes as compared with the whole embryos. The genes are shown in their chromosomal order. The genes were: in the region 19E, *Cyp6v1* through to CG12092; in 53B, CG7989 through to CG4905; and in 60D, CG4527 through to CG13579. Total RNA was extracted from 1–13-h embryos and from manually dissected adult heads and testes using Trizol reagent (Invitrogen). A total of 5 μ g RNA samples were reverse transcribed with Superscript II enzyme (Invitrogen), followed by RNase H treatment, and diluted 1:200 in water. A total of 2 μ l of diluted sample served as a template for a 25- μ l real-time PCR reaction performed in the ABI 5700 Sequence Detector, using SYBR Green chemistry (Perkin Elmer). The constitutive transcript of the *Actin 5C* gene was used as a cDNA template loading reference. Primer sequences and digital data are presented in Supplementary Information C.

included in the clusters of three genes or larger, and 45% belong to clusters of four genes or larger (Fig. 3b). The proportion of large clusters in the genome is probably even higher than shown here because, according to our data, EST profiling fails to detect about one-fifth of the testes-specific genes in *Drosophila*. Figure 2 provides specific examples of a six-gene cluster (Fig. 2a) that was detected by the EST profiling method as three singletons (Fig. 2d), and of a five-gene cluster (Fig. 2b) that was detected as a four-gene cluster (Fig. 2e).

Clustering of co-expressed homologous genes could be explained by the evolutionary history of the genomic region. The probable mechanism in this case would include local duplication(s) and divergence of amplified copies, resulting in an array of paralogues that may retain common regulatory elements. To determine the number of paralogues within the clusters of testes-specific genes, we performed protein BLAST homology searches using default BLASTP settings. Paralogues were conservatively defined as genes contained within 100 best BLASTP hits against a *Drosophila* database (expectation value 10.0). Two or more paralogues within a single cluster were found in 39 out of 239 clusters. Some of the smaller clusters consisted entirely of the paralogous genes (17 two-gene and 4 three-gene clusters) but larger clusters (11 three-gene clusters and 7 clusters of four or more genes) contained non-homologous genes along with paralogues. The relative input of local duplications could be assessed by comparing the number of genes found in clusters in excess of that expected by chance with the number of clustered paralogous genes (Fig. 3b). Whereas local duplications apparently represent the major mechanism of non-incidental formation of the two-gene clusters, they cannot account for the observed enrichment with the clusters of three or more genes. The excess of the number of genes found in clusters as compared with random gene distribution, and the trend towards large cluster size, remain statistically significant ($P \ll 0.001$) after removal of paralogues contained within the same clusters.

An unexpected finding was the unusual distribution of testes-specific genes on the X chromosome. Although about one-third of autosomal testes-specific genes were found in clusters, differing significantly from that expected for random distribution ($P \ll 0.001$), only 16% of the X-chromosomal testes-specific genes are clustered. This figure is not discernible from the 17% expected for the stochastic model ($P = 0.72$; Table 1). Similarly, the cluster size distribution is skewed towards large clusters for autosomes ($P \ll 0.001$), but not for the X chromosome ($P = 0.45$). The five-gene cluster identified in our real-time RT-PCR experiments (Fig. 2a) must represent a rare example. No similar marked differences between the X chromosome and the autosomes were found for clustering of the genes overexpressed in the head or in embryos (see below). This observation implies uniqueness in the X-chromosomal gene regulation in male development, possibly

Table 1 Clustering of tissue-specific genes in *D. melanogaster*

Gene specificity	Genome	Chr X	Chr 2	Chr 3
Testes (observed)	0.349	0.164†	0.407	0.358
Testes (random)	0.236	0.174†	0.256	0.235
Head (observed)	0.344	0.342	0.341	0.351
Head (random)	0.256	0.242	0.271	0.237
Embryo (observed)	0.335	0.310*	0.336	0.341
Embryo (random)	0.199	0.242*	0.212	0.215

The proportion of tissue-specific genes found in clusters (observed) compared with the values expected for the stochastic distribution of genes on chromosomes (random) for the indicated region of the genome is shown. The difference between the observed and the expected values is highly significant ($P \ll 0.001$) in all cases except where noted.

*Significant difference ($P = 0.016$) for X-linked embryonic genes.

†No significant difference ($P = 0.72$) for X-linked testes-specific genes.

owing to the dosage compensation⁵ that involves extensive acetylation of the X chromatin in males^{6–8}.

We observed clustering of genes that are co-expressed in spermatogenesis. Clusters of genes co-regulated during circadian cycles have also been reported in *Drosophila*⁴. To determine whether this is a general trait of *Drosophila* genome architecture, we performed EST-based analysis of the distribution of head-specific genes. The pattern was similar to that observed for testes-specific genes. The proportion of clustered genes was substantially higher than expected by chance (Table 1), with a significant trend towards larger clusters ($P \ll 0.001$). Real-time RT-PCR analysis confirmed the EST-based prediction of head-specific gene clustering in the genomic region 55C by identification of a nine-gene cluster (CG15068, CG15065, CG16836, CG16844, *Im2*, CG15067, CG18107, CG18108 and CG15066; see Supplementary Information C). Cognate results were obtained for genes overexpressed in the embryo, with the proportion of genes found in clusters significantly exceeding that expected for random distribution (Table 1), and with an enrichment towards larger clusters ($P \ll 0.001$).

Our analysis implies that one-third of tissue-specific genes are clustered in the *D. melanogaster* genome according to their transcription patterns. A significant trend towards large clusters is discernible. About 45% of genes found in clusters in excess of that expected by chance belong to three-gene clusters, and another 45% are organized in clusters of four or more genes. Large groups of co-expressed genes in *Drosophila*⁹ seem to pertain to the same phenomenon. Large clusters of non-homologous, co-expressed genes have also been found in nematodes¹⁰ and in mammals^{3,11–13}, indicating that the trend is common in higher eukaryotes. Conversely, clusters of co-expressed genes found in yeast comprise mostly gene pairs¹. This bias may reflect the prevalent use of the mechanisms that operate with extensive chromosomal domains, such as the chromatin potentiation^{11,14} in higher eukaryotes, in contrast with the idea of widespread co-regulation through shared transcriptional regulatory elements in lower eukaryotes¹.

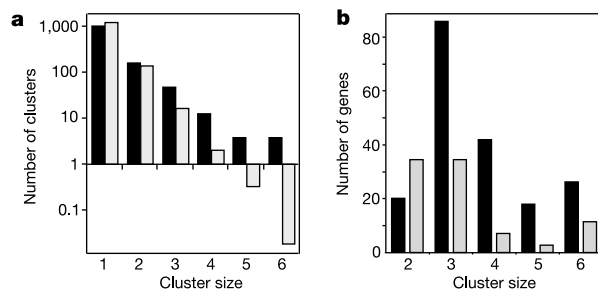


Figure 3 The *D. melanogaster* genome is enriched with large clusters of non-homologous testes-specific genes. Cluster size is the number of genes per cluster. **a**, The height of the bars represents the number of testes-specific gene clusters of corresponding size revealed by EST profiling (black) or estimated for the stochastic distribution of testes-

specific genes (grey). **b**, The distribution of the genes found in clusters of different sizes in excess of that expected by chance (black), and the number of paralogous genes detected within clusters (grey).

Potential of chromatin is a prerequisite for transcriptional activation, and has been linked to the cooperative activity of ATP-dependent chromatin remodelling complexes and histone acetyl transferases^{15–18}. Reported interference between the chromatin remodelling complexes and acetylation of the histone H4 on the X chromosome in males induced by the Male-Specific Lethal (MSL)-dependent dosage compensation machinery¹⁹ may be responsible for the observed scarcity of testes-specific gene clusters on the X chromosome, as well as for the lower ratio of male-specific genes on the X chromosome as compared with autosomes^{20,21}. The region around the gene *runt* where we did find clusters of testes-specific genes (Fig. 1a) is the only known segment of the X chromosome devoid of the MSL-induced H4 acetylation²². These data imply an evolutionary exodus of male-specific genes from the X chromosome to autosomes in flies to avoid interference with the dosage compensation mechanism, consistent with the presented concept of widespread chromatin-mediated transcriptional co-regulation of tissue-specific genes. □

Received 3 May; accepted 7 October 2002; doi:10.1038/nature01216.

- Cohen, B. A., Mitra, R. D., Hughes, J. D. & Church, G. M. A computational analysis of whole-genome expression data reveals chromosomal domains of gene expression. *Nature Genet.* **26**, 183–186 (2001).
- Cho, R. J. *et al.* A genome-wide transcriptional analysis of the mitotic cell cycle. *Mol. Cell* **2**, 65–73 (1998).
- Lercher, M. J., Urrutia, A. O. & Hurst, L. D. Clustering of housekeeping genes provides a unified model of gene order in the human genome. *Nature Genet.* **31**, 180–183 (2002).
- Ueda, H. R. *et al.* Genome-wide transcriptional orchestration of circadian rhythms in *Drosophila*. *J. Biol. Chem.* **277**, 14048–14052 (2002).
- Meller, V. H. Dosage compensation: making 1X equal 2X. *Trends Cell Biol.* **10**, 54–59 (2000).
- Akhtar, A. & Becker, P. B. Activation of transcription through histone H4 acetylation by MOF, an acetyltransferase essential for dosage compensation in *Drosophila*. *Mol. Cell* **5**, 367–375 (2000).
- Smith, E. R. *et al.* The *Drosophila* MSL complex acetylates histone H4 at lysine 16, a chromatin modification linked to dosage compensation. *Mol. Cell Biol.* **20**, 312–318 (2000).
- Bhadra, U., Pal-Bhadra, M. & Birchler, J. A. Role of the male specific lethal (*msl*) genes in modifying the effects of sex chromosomal dosage in *Drosophila*. *Genetics* **152**, 249–268 (1999).
- Spellman, P. T. & Rubin, G. M. Evidence for large domains of similarly expressed genes in the *Drosophila* genome. *J. Biol.* **1**, 5 (2002).
- Roy, P. J., Stuart, J. M., Lund, J. & Kim, S. K. Chromosomal clustering of muscle-specific genes in *Caenorhabditis elegans*. *Nature* **418**, 975–979 (2002).
- Kramer, J. A., McCarrey, J. R., Djakiew, D. & Krawetz, S. A. Differentiation: the selective potentiation of chromatin domains. *Development* **125**, 4749–4755 (1998).
- Wynn, S. L. *et al.* Organization and conservation of the *GART/SON/DONSON* locus in mouse and human genomes. *Genomics* **68**, 57–62 (2000).
- Soury, E. *et al.* Chromosomal assignment of mammalian genes with an acute inflammation-regulated expression in liver. *Immunogenetics* **53**, 634–642 (2001).
- Kramer, J. A., McCarrey, J. R., Djakiew, D. & Krawetz, S. A. Human spermatogenesis as a model to examine gene potentiation. *Mol. Reprod. Dev.* **56**, 254–258 (2000).
- Schubeler, D. *et al.* Nuclear localization and histone acetylation: a pathway for chromatin opening and transcriptional activation of the human β -globin locus. *Genes Dev.* **14**, 940–950 (2000).
- Elefant, F., Cooke, N. E. & Liehaber, S. A. Targeted recruitment of histone acetyltransferase activity to a locus control region. *J. Biol. Chem.* **275**, 13827–13834 (2000).
- Brown, C. E., Lechner, T., Howe, L. A. & Workman, J. L. The many HATs of transcription coactivators. *Trends Biochem. Sci.* **25**, 15–19 (2000).
- Kingston, R. E. & Narlikar, G. J. ATP-dependent remodeling and acetylation as regulators of chromatin fluidity. *Genes Dev.* **13**, 2339–2352 (1999).
- Corona, D. F., Clapier, C. R., Becker, P. B. & Tamkun, J. W. Modulation of ISWI function by site-specific histone acetylation. *EMBO Rep.* **3**, 242–247 (2002).
- Meiklejohn, C. D., Parsch, J., Ranz, J. M. & Hartl, D. L. *43rd Ann. Drosophila Research Conference, Abstract 112 40* (The Genetics Society of America, Bethesda, 2002).
- Parisi, M. J. *et al.* *43rd Ann. Drosophila Research Conference, Abstract 46 17* (The Genetics Society of America, Bethesda, 2002).
- Smith, E. R., Allis, C. D. & Lucchesi, J. C. Linking global histone acetylation to the transcription enhancement of X-chromosomal genes in *Drosophila* males. *J. Biol. Chem.* **276**, 31483–31486 (2001).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements This work was supported by a grant from the US Public Health Service. We thank D. Albertini, E. Gibney, M. Nurminkaya and V. Shkadov for their comments and careful reading of the manuscript.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to D.L.N. (e-mail: dmitry.nurminkaya@tufts.edu).

The mechanosensory protein MEC-6 is a subunit of the *C. elegans* touch-cell degenerin channel

Dattananda S. Chelur*, Glen G. Ernstrom*, Miriam B. Goodman*†, C. Andrea Yao*, Lei Chen*, Robert O' Hagan* & Martin Chalfie*

* Department of Biological Sciences, Columbia University, New York, New York 10027, USA

† Department of Molecular and Cellular Physiology, Stanford University School of Medicine, Stanford, California 94305-5345, USA

Mechanosensory transduction in touch receptor neurons is believed to be mediated by DEG/ENaC (degenerin/epithelial Na⁺ channel) proteins in nematodes and mammals¹. In the nematode *Caenorhabditis elegans*, gain-of-function mutations in the degenerin genes *mec-4* and *mec-10* (denoted *mec-4(d)* and *mec-10(d)*, respectively) cause degeneration of the touch cells^{2,3}. This phenotype is completely suppressed by mutation in a third gene, *mec-6* (refs 3, 4), that is needed for touch sensitivity. This last gene is also required for the function of other degenerins^{4–8}. Here we show that *mec-6* encodes a single-pass membrane-spanning protein with limited similarity to paraoxonases, which are implicated in human coronary heart disease⁹. This gene is expressed in muscle cells and in many neurons, including the six touch receptor neurons. MEC-6 increases amiloride-sensitive Na⁺ currents produced by MEC-4(d)/MEC-10(d) by ~30-fold, and functions synergistically with MEC-2 (a stomatin-like protein¹⁰ that regulates MEC-4(d)/MEC-10(d) channel activity¹¹) to increase the currents by 200-fold. MEC-6 physically interacts with all three channel proteins. *In vivo*, MEC-6 co-localizes with MEC-4, and is required for punctate MEC-4 expression along touch-neuron processes. We propose that MEC-6 is a part of the degenerin channel complex that may mediate mechanotransduction in touch cells.

We cloned *mec-6* by refining its genetic map position, identifying a rescuing cosmid in the region, and identifying a predicted gene (W02D3.3) that was partly or completely deleted in three *mec-6* strains (Fig. 1a). We confirmed this identification by transformation rescue of *mec-6* mutants with DNA for W02D3.3. *mec-6* encodes a messenger RNA of approximately 1.2 kilobases (1.2 kb) on northern blots (Fig. 1b). The *mec-6(u450)* mutation is a null mutation, as it lacks the promoter and the first four exons of the coding sequence (Fig. 1a) and does not produce any detectable mRNA (Fig. 1b). All eleven *mec-6* strains that we sequenced had defects in the same gene (Fig. 1c).

mec-6 complementary DNA encodes a polypeptide of 377 amino acids (Fig. 1c) that is 25% identical and 45% similar over a stretch of 250 amino acids at its carboxy terminus to vertebrate paraoxonase/arylesterase. Humans and mice have three genes encoding similar paraoxonase (PON) proteins: *PON1*, *PON2* and *PON3* (ref. 12). *PON1* and *PON3* are HDL (high-density lipoprotein)-associated serum proteins¹³, whereas *PON2* is ubiquitously expressed and is localized to the plasma membrane¹⁴. All three PONs prevent the lipid oxidation of LDL particles, and so are thought to have a role in atherosclerosis and coronary heart disease^{9,13,14}. The physiological function of the PONs, however, is not known. MEC-6 and four similar *C. elegans* proteins contain a 15-amino-acid sequence in their amino termini (Fig. 1c) that is absent in the vertebrate PONs. The relationship of MEC-6, the PONs and other similar nematode proteins is given in Supplementary Information.

Many cells, including the six touch receptor neurons, expressed green fluorescent protein (GFP) and β -galactosidase (LacZ) from *mec-6* promoter and protein fusions (Fig. 1d and data not shown).

These fusions were expressed primarily in neurons (the six touch receptors, ventral cord motor neurons, HSN, PVD, PVC, IL1, and several other unidentified neurons near the nerve ring, in the anal ganglion and in the male tail sensory rays), muscles (the body wall, vulval, intestinal, anal depressor and sphincter muscles) and the excretory canal. Despite the expression of *mec-6* in many cell types, expression of full-length *mec-6::3Xflag* from the touch-cell-specific *mec-18* promoter produced touch-sensitive animals. The widespread expression pattern of *mec-6* and its overlap with the expression pattern of the degenerin genes *deg-1*, *unc-8*, *del-1* and *unc-105* (refs 6–8) are consistent with the suppression by *mec-6* of degenerations caused by the ectopic expression of *mec-4(d)* in several different cell types¹⁵ and of gain-of-function mutations of other degenerin genes^{4–8}. *mec-6* may be required for the function of many of the 20 or more degenerin proteins encoded by the *C. elegans* genome¹⁶.

MEC-6 contains a predicted transmembrane domain and two potential glycosylation sites (Fig. 1c), suggesting that MEC-6 is a membrane protein with a very short cytoplasmic N terminus and a large extracellular C terminus. We tested this topological model by fusing *mec-6* to *lacZ*, because cytoplasmic, but not extracellular, β -galactosidase has enzymatic activity (see, for example, ref. 17). A full-length *mec-6::lacZ* fusion produced no β -galactosidase activity unless DNA encoding a synthetic transmembrane domain was added (Fig. 1e). The transmembrane domain-dependent changes in *lacZ* expression indicate that MEC-6 is a transmembrane protein with its C terminus situated extracellularly. Additional support for the transmembrane topology of MEC-6 is presented in Supplementary Information.

We previously showed that the expression of MEC-4(d) or MEC-4(d)/MEC-10(d) results in degenerin channel activity in *Xenopus* oocytes¹¹. The ability of *mec-6* mutations to block touch-cell degeneration caused by *mec-4(d)* and *mec-10(d)* suggested that wild-type MEC-6 is critical for degenerin channel activity. To test this possibility, we co-expressed MEC-6 with the mutant degenerins MEC-4(d) and MEC-10(d) in *Xenopus* oocytes. In these experiments, we used less *mec-6* RNA (1 ng per oocyte) than for the other proteins because oocytes expressing MEC-6 (10 ng per oocyte) contained an inwardly rectified and amiloride-resistant current (Supplementary Information).

Injection of *mec-6* RNA at 1 ng per oocyte not only minimized the production of the amiloride-resistant currents, but also maximized the amiloride-sensitive currents of MEC-4(d)/MEC-10(d) channels (data not shown). On average, the addition of *mec-6* RNA increased the size of the amiloride-sensitive currents 24-fold at -85 mV (Fig. 2a). MEC-10(d) is not required for this effect, indicating that MEC-4(d) and MEC-6 alone are sufficient to form robust channels, as inferred previously from the observation that wild-type *mec-6* was sufficient for the degenerations caused by ectopically expressed *mec-4(d)*¹⁵. Injecting MEC-6 and MEC-10(d) RNA together produced currents that were indistinguishable from water-injected control cells ($I(-85 \text{ mV}) = -0.03 \pm 0.05 \mu\text{A}$, $n = 11$). Co-expression of MEC-6 with MEC-4(d)/MEC-10(d) resulted in a sodium-selective current ($P_{\text{Li}} : P_{\text{Na}} : P_{\text{K}} = 1.70 : 1 : 0.08$) that was blocked by the pore-blocker amiloride. The apparent affinity for amiloride (K_i') was $0.14 \pm 0.02 \mu\text{M}$, $n = 13$ at -100 mV. MEC-6 is present at the oocyte cell surface, and MEC-6 increases average current size without increasing the amount of MEC-4(d) or MEC-10(d) protein detected in the plasma membrane (Fig. 2b). MEC-6 is, therefore, likely to increase macroscopic current by increasing the average current carried by single channels. This change could be accomplished by increasing single-channel conductance, mean open time, or open probability.

We previously showed that MEC-2 increases MEC-4(d) and MEC-4(d)/MEC-10(d) current amplitude by ~ 45 -fold¹¹. Here we show that the effects of MEC-2 and MEC-6 on current amplitude are synergistic; together they increase MEC-4(d) currents more than

400-fold. This finding suggests that MEC-2 and MEC-6 increase current through different mechanisms. Co-expression of all four proteins produced a smaller (199-fold) increase that was, nonetheless, greater than expected from a simple additive process (Fig. 2a). The expression of the four proteins resulted in a sodium-selective current ($P_{\text{Li}} : P_{\text{Na}} : P_{\text{K}} = 1.54 : 1 : 0.12$, $n = 26$) that was blocked by amiloride. The apparent affinity for amiloride was $0.27 \pm 0.03 \mu\text{M}$ ($n = 15$) at -100 mV, a value similar to that reported for rat $\alpha\beta\gamma\text{ENaC}$ channel ($K_i' = 0.104 \mu\text{M}$ at -100 mV)¹⁸.

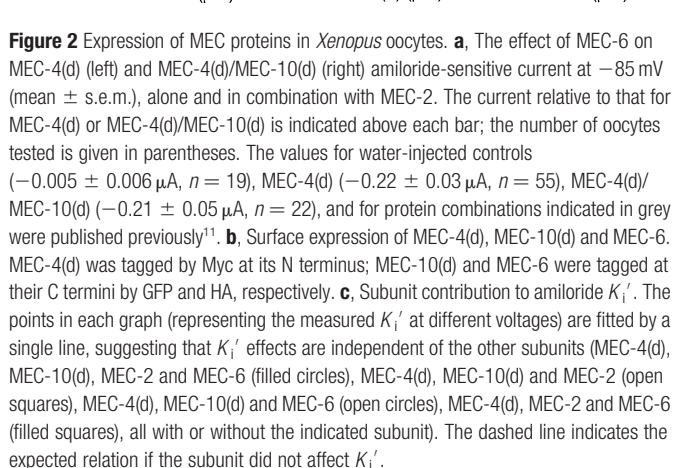
In contrast to the relatively small amiloride-sensitive current produced by MEC-4(d)/MEC-10(d), co-expression with MEC-6, MEC-2 or both resulted in robust amiloride-sensitive currents. These substantial currents allowed us to analyse the contribution of each subunit to amiloride affinity and relative ion permeability. Neither the K_i' nor the voltage dependence of amiloride blockade (overall $\delta_{+\text{MEC-6}}/\delta_{-\text{MEC-6}} = 0.96$) was significantly altered by the addition of MEC-6 (Supplementary Information and Fig. 2c). MEC-6 also does not seem to introduce additional classes of binding sites, as dose-response curves in the presence of MEC-6 were effectively described by a single-site binding isotherm (data not shown). Thus, MEC-6 is unlikely to contribute to the amiloride-binding site. By contrast, MEC-10(d) reduces amiloride K_i' of the



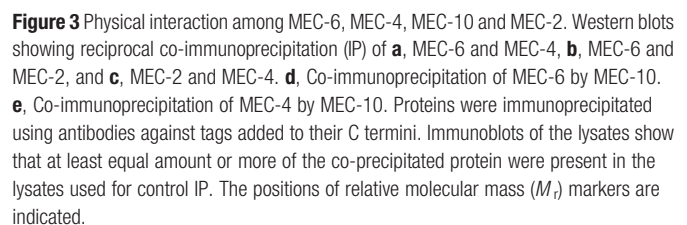
Figure 1 Characterization of *mec-6*. **a**, The 7.2-kb rescuing *mec-6* genomic DNA, with exons (boxes), introns (line) and deletions (bars) indicated. **b**, Northern blots of wild-type and *mec-6* mutants probed with *mec-6* DNA and re-probed with initiation factor-1, *inf-1*. **c**, Deduced MEC-6 protein sequence. The predicted transmembrane domain (underline), myristoylation site (green), glycosylation sites (red) and the *C. elegans*-specific domain (boxed) are indicated. Missense (letter), nonsense (asterisk), frameshift (arrow) and splice-site (inverted triangle) mutations are indicated above the affected amino acid for each mutant allele (in parentheses). **d**, Expression of full-length *mec-6::yfp* fusions in an L3 larva (top) and adults (bottom, left to right: body wall muscle, vulval muscle and the excretory canal). **e**, Expression of full-length *mec-6::lacZ* fusions with or without a synthetic transmembrane domain in an L3 larva. Arrows indicate touch cells in **d** and **e**. In the schematic drawings, MEC-6 with its transmembrane domain (thicker bar) is black, LacZ is blue, and the synthetic transmembrane domain is red.

We also examined the effect of the various subunits on ion selectivity by measuring the change in reversal potential caused by substituting K^+ or Li^+ for Na^+ . As with MEC-2 (ref. 11), the presence of MEC-6 lowered the P_{Li}/P_{Na} ratio of the MEC-4(d)/

Responsible	Percentage
Current administration	85
Previous administrations	15



To study protein–protein interactions, we tagged these proteins with different epitopes at their C termini and expressed them in Chinese hamster ovary (CHO) cells. The tagged proteins retained their normal function, as they rescued touch insensitivity in *C. elegans* (data not shown). MEC-6 reciprocally co-immunoprecipitated MEC-4 and MEC-2, and MEC-2 and MEC-4 co-precipitated each other (Fig. 3a–c). MEC-10 co-precipitated MEC-4 and MEC-6 (Fig. 3d, e); the reverse precipitation could not be done because of low expression of MEC-10 in the CHO cells (it could not be detected in western blots of the lysate without prior immunoprecipitation). These interactions seem specific, as none of the proteins co-precipitated membrane localized cyan fluorescent protein (CFP), an unrelated protein. These protein–protein interactions among MEC-2, MEC-4, MEC-6 and MEC-10 provide a biochemical basis for the functional interactions in oocytes as well as the previously established genetic interactions^{3,4,19}. MEC-6 is also likely to interact



similarly with other degenerins, such as the products of *deg-1*, *unc-8* and *unc-105*, whose gain-of-function phenotypes are either completely or partially suppressed by *mec-6* (refs 4–7; D.S.C. and M.C., unpublished data).

MEC-4 and MEC-6 co-localize to discrete puncta along the length of touch-cell axons, providing further evidence that these proteins exist as a complex *in vivo*. A full-length *mec-4::yfp* protein fusion, but not a *P_{mec-4}cfp* promoter fusion, was expressed as distinct dots along the entire length of the six touch-cell processes (Fig. 4a). Full-length 3XFlag-tagged MEC-6 was also present as puncta that overlapped those of MEC-4::YFP (Fig. 4b). *mec-6* expression was punctuate in CHO cells and in *C. elegans* muscle cells (Fig. 1d), where it overlapped with the puncta produced by MEC-4::YFP expressed ectopically using a muscle-specific promoter (D.S.C. and M.C., unpublished data).

The *mec-6* gene is required for the punctate distribution of MEC-4. When *mec-4::yfp* and *P_{mec-4}cfp* were injected into *mec-6(u3)* animals, the expression of the *P_{mec-4}cfp* fusion was unchanged, but the punctate expression from *mec-4::yfp* was undetectable (Fig. 4c). Because *mec-6* mutations affected this punctate pattern and not the expression from the promoter fusion, MEC-6 does not regulate *mec-4* transcription (a result confirmed by northern blotting; data not shown). MEC-6 thus appears to affect the localization and/or stability of MEC-4. The effect of *mec-6* on localization/stability of MEC-4 is specific to the touch cells, suggesting that additional factor(s) present in the touch cells may be required for this function. We could not determine if *mec-10* had similar punctate expression or co-localized with the other two

proteins, because the expression of full-length *mec-10::gfp* or *mec-10::myc* fusions was very weak in the touch-cell processes (data not shown).

The punctate expression of MEC-6 in the touch cells, and in other cells, is reminiscent of the localization pattern of proteins that are associated with membrane microdomains like lipid rafts or caveolae. Several signal transduction molecules and channel proteins have been recently shown to cluster at such specialized microdomains²⁰. Indeed, there is biochemical evidence²¹ for ENaC being localized to lipid rafts. MEC-6 may be necessary for clustering of the degenerin channels to specific membrane domains needed for mechanosensory transduction. The similarity of MEC-6 to PON proteins suggests that MEC-6 might, like the PONs, interact with lipids and this interaction might, in turn, be crucial for the localization of the degenerin channel complexes to the specialized membrane domains.

The cloning of *mec-6* has helped us to reconstitute the touch-cell degenerin channel complex that is postulated to mediate mechanosensory transduction. Interaction and co-localization of MEC-6 with degenerin proteins provides a biochemical basis for understanding the genetic interactions among these genes, and supports the idea that all four proteins form a single ion channel complex. As MEC-6 has a large extracellular domain like MEC-4 and MEC-10, the interaction among these proteins and, perhaps, with the extracellular matrix components needed for touch sensitivity (for example, MEC-1, MEC-5 and MEC-9) might occur via the extracellular domain. The interaction with MEC-2 might occur via the short N-terminal region or the transmembrane domain. Such interactions may be important in regulating degenerin channel activity. The similarity of MEC-6 to PONs also suggests that the vertebrate PONs, which are expressed in many tissues (ref. 14, and as revealed by the representation in the expressed-sequence tag (EST) division of GenBank), might have a previously unrecognized function: to interact with DEG/ENaC channel proteins and stomatin-like proteins to regulate channel activity. For example, PON ESTs (GenBank accession numbers AI986572 and BG742085) have been identified from rat dorsal root ganglia and skin, tissues that also express BNC1 (refs 22, 23), a DEG/ENaC channel that has been implicated in mechanotransduction²³. These findings raise the possibility that at least some of the vertebrate PONs might be needed for mechanosensory transduction. □

Methods

Mapping and cloning of *mec-6*

mec-6 was cloned by refining its genetic map position, testing cosmids from the region for phenotypic rescue of touch insensitivity by germline transformation, and testing for deletion of characteristic polymerase chain reaction (PCR) fragments from predicted genes in the rescuing cosmid in *mec-6* mutants. Details are given in Supplementary Information.

Molecular characterization of *mec-6*

We screened a *C. elegans* embryonic cDNA library²⁴ with a 3.5-kb genomic DNA fragment containing the entire coding sequence of *mec-6*. Of 22 individual clones, 11 contained a 1.2-kb insert and 11 contained a 1.4-kb insert. Two clones for each size of transcript were sequenced (Sequenase 2.0 kit; US Biochemical). Both transcripts were SL1 trans-spliced and contained the same coding sequence; they differed in the length of the 3'-untranslated region. These cDNAs (GenBank accession number AF295927) encoded a protein that is five residues shorter than the predicted protein W02D3.3 because the splicing signal at the end of the fifth exon was predicted incorrectly. We identified mutations in *mec-6* alleles by sequencing genomic DNA amplified from single worms using PCR.

Fluorescent protein and lacZ fusions

All fluorescent protein and lacZ fusions were created using the Fire lab vector kit (a gift from A. Fire). *mec-4* promoter fusions and full-length protein fusions with two coloured variants of GFP (CFP and YFP²⁵) were created at the 16th codon and at the end of coding sequences of *mec-4* genomic DNA, respectively. Similarly, in-frame *mec-6* promoter fusions were generated at the 11th codon and the full-length rescuing fusion was created by addition next to the last codon. The initiator ATG codon of *lacZ* or *yfp* in the Fire vector was either removed or changed to ATT before creating the rescuing full-length *mec-6* fusions to eliminate the strong gut expression from an internal promoter within the *mec-6* sequence. For the transmembrane construct, a synthetic transmembrane domain (*KpnI* fragment) from pPD34.110 (ref. 26) was inserted between the *mec-6* and *lacZ* sequences.

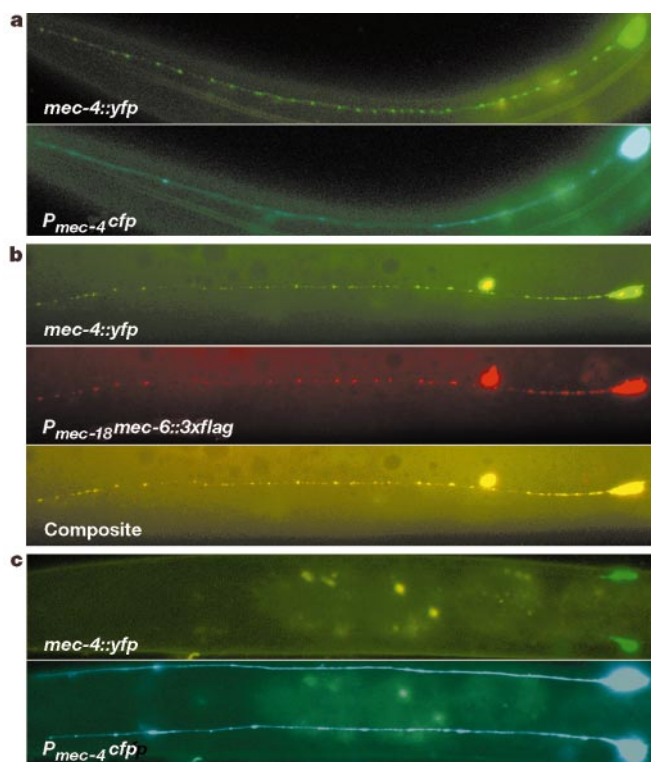


Figure 4 *mec-6* co-localizes with MEC-4 in a punctate pattern in the touch cells. **a**, Expression of *mec-4::yfp* and *P_{mec-4}cfp* in a wild-type adult. YFP and CFP fluorescence was monitored in the same nematode using appropriate filter sets. **b**, Co-localization of MEC-4 and MEC-6. Expression of *P_{mec-18}mec-6::3xflag* was detected by immunostaining with anti-Flag antibody and rhodamine-conjugated secondary antibody. The full-length *mec-4::yfp* was also expressed in the same animal. The two images were merged to form the composite; overlap is seen as yellow. **c**, Expression of *mec-4::yfp* and *P_{mec-4}cfp* in a *mec-6(u3)* animal.

For antibody staining, the 3XFlag epitope (from p3XFlag-CMV-14; Sigma), fused at the end of coding sequences of *mec-6* genomic DNA, was expressed from the strong, touch-cell-specific *mec-18* promoter (G. Gu and M.C., unpublished data).

Expression in CHO cells, immunoprecipitation and immunostaining

Constructs encoding C-terminally tagged proteins were expressed in CHO-K1 cells. Wild-type *mec-4* (ref. 17) and *mec-10* (ref. 3) cDNAs were cloned in the vectors p3XFlag-CMV-14 (Sigma) and pCDNA3.1/Myc-His(+) (Invitrogen), respectively, and grown in *Escherichia coli* strain SMC4 (ref. 11). Wild-type *mec-2* cDNA was cloned in pCDNA3.1/Myc-His(+) and a PCR primer containing the entire sequence of the nine-amino-acid haemagglutinin (HA) peptide was used to add the HA-tag at the end of the coding sequence of *mec-6* cDNA in pCDNA3.1. *mec-6* cDNA was inserted into p3XFlag-CMV-14 to make a *mec-6::3XFlag* construct. These other constructs were propagated in XL-2 Blue (Stratagene). CHO cells were cultured in F12 nutrient medium with 10% bovine fetal serum. Cells, transiently transfected using Lipofectamine2000 (Invitrogen), were collected after 24 h and lysed in IP buffer (20 mM Tris, pH 7.5, 150 mM NaCl, 10% glycerol, 0.5% Triton X-100 and protease inhibitor cocktail (Roche)). The lysate was incubated with agarose-conjugated antibodies (Santa Cruz Biotechnology) in the same buffer for 16 h at 4 °C, and the precipitated proteins were analysed on western blot using peroxidase-conjugated antibodies (Santa Cruz Biotechnology) and the ECL plus chemiluminescence kit (Amersham). Membrane-localized CFP, used as a control in the immunoprecipitation reaction, was expressed from pECFP-Mem (Clontech). For surface immunostaining, the MEC-6::HA transfected cells were fixed in 4% paraformaldehyde in PBS and incubated with anti-HA rat monoclonal antibody (Roche) followed by Texas-red-conjugated anti-rat secondary antibodies.

Oocyte electrophysiology and surface expression

Construction of *mec-2*, *mec-4(d)* and *mec-10(d)* plasmids in oocyte expression vector pGEM-HE, *mec-4(d)* and *mec-10(d)* propagation in SMC4, electrophysiological methods, and analysis of the surface expressed proteins have been described¹¹. Wild-type *mec-6* cDNA was cloned in pGEM-HE and the G235R mutation of *mec-6(u3)* was introduced by *in vitro* mutagenesis. Oocytes were collected from female *Xenopus laevis* frogs (Xenopus I or Nasco) and injected with 10 ng of capped RNA encoding MEC-4(d), MEC-10(d) and MEC-2, and 1 ng of capped RNA encoding MEC-6. Oocytes were maintained in L-15 oocyte medium containing 100 µg ml⁻¹ gentamicin (Cell & Molecular Technologies) and 300 µM amiloride (Sigma). Membrane current was measured 4–10 d after RNA injection. Relative ion permeability and amiloride K_i⁺ were determined as described¹¹. Relative ion permeabilities were measured from the reversal potentials of amiloride difference currents except for MEC-4(d) + MEC-2, which was measured from the reversals of total current. The K_i⁺ and ion selectivities of oocytes expressing both MEC-2 and MEC-6 with either MEC-4(d) or MEC-4(d) + MEC-10(d) were measured 2–4 d after RNA injection.

Received 7 August; accepted 20 August 2002; doi:10.1038/nature01205.

- Ernstrom, G. G. & Chalfie, M. Genetics of sensory mechanotransduction. *Annu. Rev. Genet.* (in the press).
- Driscoll, M. & Chalfie, M. The *mec-4* gene is a member of a family of *Caenorhabditis elegans* genes that can mutate to induce neuronal degeneration. *Nature* **349**, 588–593 (1991).
- Huang, M. & Chalfie, M. Gene interactions affecting mechanosensory transduction in *Caenorhabditis elegans*. *Nature* **367**, 467–470 (1994).
- Chalfie, M. & Wolinsky, E. The identification and suppression of inherited neurodegeneration in *Caenorhabditis elegans*. *Nature* **345**, 410–416 (1990).
- Shreffler, W., Magardino, T., Shekdar, K. & Wolinsky, E. The *unc-8* and *sup-40* genes regulate ion channel function in *Caenorhabditis elegans* motoneurons. *Genetics* **139**, 1261–1272 (1995).
- García-Añoveros, J., Ma, C. & Chalfie, M. Regulation of *Caenorhabditis elegans* degeneration proteins by a putative extracellular domain. *Curr. Biol.* **5**, 441–448 (1995).
- Liu, J., Schrank, B. & Waterston, R. H. Interaction between a putative mechanosensory membrane channel and a collagen. *Science* **273**, 361–364 (1996).
- Tavernarakis, N., Shreffler, W., Wang, S. & Driscoll, M. *unc-8*, a DEG/ENAC family member, encodes a subunit of a candidate mechanically gated channel that modulates *C. elegans* locomotion. *Neuron* **18**, 107–119 (1997).
- Durrington, P. N., Mackness, B. & Mackness, M. I. Paraoxonase and atherosclerosis. *Arterioscler. Thromb. Vasc. Biol.* **21**, 473–480 (2001).
- Huang, M., Gu, G., Ferguson, E. L. & Chalfie, M. A stomatin-like protein necessary for mechanotransduction in *C. elegans*. *Nature* **378**, 292–295 (1995).
- Goodman, M. B. *et al.* MEC-2 regulates *C. elegans* DEG/ENAC channels needed for mechanosensation. *Nature* **415**, 1039–1042 (2002).
- Primo-Parmo, S. L., Sorenson, R. C., Teiber, J. & La Du, B. N. The human serum paraoxonase/arylesterase gene (PON1) is one member of a multigene family. *Genomics* **33**, 498–507 (1996).
- Draganov, D. I., Stetson, P. L., Watson, C. E., Billecke, S. S. & La Du, B. N. Rabbit serum paraoxonase 3 (PON3) is a high density lipoprotein-associated lactonase and protects low density lipoprotein against oxidation. *J. Biol. Chem.* **275**, 33435–33442 (2000).
- Ng, C. J. *et al.* Paraoxonase-2 is a ubiquitously expressed protein with antioxidant properties, and is capable of preventing cell-mediated oxidative modification of low-density lipoprotein. *J. Biol. Chem.* **276**, 44444–44449 (2001).
- Harbinder, S. *et al.* Genetically targeted cell disruption in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **94**, 13128–13133 (1997).
- The *C. elegans* Sequencing Consortium. Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* **282**, 2012–2018 (1998).
- Lai, C. C., Hong, K., Kinnell, M., Chalfie, M. & Driscoll, M. Sequence and transmembrane topology of MEC-4, an ion channel subunit required for mechanotransduction in *Caenorhabditis elegans*. *J. Cell. Biol.* **133**, 1071–1081 (1996).
- Canessa, C. M. *et al.* Amiloride-sensitive epithelial Na⁺ channel is made of three homologous subunits. *Nature* **367**, 463–467 (1994).

- Gu, G., Caldwell, G. A. & Chalfie, M. Genetic interactions affecting touch sensitivity in *Caenorhabditis elegans*. *Proc. Natl Acad. Sci. USA* **93**, 6577–6582 (1996).
- Galbiati, F., Razani, B. & Lisanti, M. P. Emerging themes in lipid rafts and caveolae. *Cell* **106**, 403–411 (2001).
- Hill, W. G., An, B. & Johnson, A. P. Endogenously expressed epithelial sodium channel is present in lipid rafts in A6 cells. *J. Biol. Chem.* **277**, 33541–33544 (2002).
- Price, M. P. *et al.* The mammalian sodium channel BNC1 is required for normal touch sensation. *Nature* **407**, 1007–1011 (2000).
- García-Añoveros, J., Samad, T. A., Zuvela-Jelaska, L., Woolf, C. J. & Corey, D. P. Transport and localization of the DEG/ENAC ion channel BNC1α to peripheral mechanosensory terminals of dorsal root ganglia neurons. *J. Neurosci.* **21**, 2678–2686 (2001).
- Okkema, P. G. & Fire, A. The *Caenorhabditis elegans* NK-2 class homeoprotein CEH-22 is involved in combinatorial activation of gene expression in pharyngeal muscle. *Development* **120**, 2175–2186 (1994).
- Fire, A., Harrison, S. W. & Dixon, D. A modular set of *lacZ* fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189–198 (1990).
- Miller, D. M. III *et al.* Two-colour GFP expression system for *C. elegans*. *Biotechniques* **26**, 914–921 (1999).

Supplementary Information accompanies the paper on Nature's website
(<http://www.nature.com/nature>).

Acknowledgements We thank J. Kong for technical assistance in isolating *mec-6* cDNAs. This work was supported by the National Institutes of Health (M.C.), a Human Frontiers Science Program postdoctoral fellowship (D.S.C.), and a postdoctoral fellowship from the National Institute of Deafness and other Communication Disorders (M.B.G.).

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.C.
(e-mail: mc21@columbia.edu).

Role for Slimb in the degradation of *Drosophila* Period protein phosphorylated by Doubletime

Hyuk Wan Ko⁺, Jin Jiang[†] & Isaac Edery[‡]

⁺ Graduate Program in Physiology and Neurobiology, and [‡] Department of Molecular Biology and Biochemistry, Rutgers University, Center for Advanced Biotechnology and Medicine, 679 Hoes Lane, Piscataway, New Jersey 08854, USA
[†] Center for Developmental Biology and Department of Pharmacology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA

Protein phosphorylation has a key role in modulating the stabilities of circadian clock proteins in a manner specific to the time of day¹. A conserved feature of animal clocks is that Period (Per) proteins undergo daily rhythms in phosphorylation and levels^{2,3}, events that are crucial for normal clock progression^{4–7}. Casein kinase Iε (CKIε) has a prominent role in regulating the phosphorylation and abundance of Per proteins in animals⁸. This was first shown in *Drosophila* with the characterization of Doubletime (Dbt), a homologue of vertebrate casein kinase Iε^{4,6}. However, it is not clear how Dbt regulates the levels of Per. Here we show, using a cell culture system, that Dbt promotes the progressive phosphorylation of Per, leading to the rapid degradation of hyperphosphorylated isoforms by the ubiquitin–proteasome pathway. Slimb, an F-box/WD40-repeat protein functioning in the ubiquitin–proteasome pathway^{9,10} interacts preferentially with phosphorylated Per and stimulates its degradation. Overexpression of *slimb* or expression in clock cells of a dominant-negative version of *slimb* disrupts normal rhythmic activity in flies. Our findings suggest that hyperphosphorylated Per is targeted to the proteasome by interactions with Slimb.

There have been no reports of Dbt-dependent phosphorylation of Per in a heterologous system, which restricts the ability to address

mechanistic issues. To circumvent this limitation we sought to establish a simplified system by using transfected cultured *Drosophila* S2 cells. Expression of *per* was controlled from a constitutive promoter (pAct-*per*), whereas the copper inducible metallothionein (pMT) promoter regulated *dbt* expression (pMT-*dbt*). In some cases expressed proteins contained either the V5 or the haemagglutinin (HA) epitope tag to facilitate detection.

Induction of *dbt-V5* led to progressive decreases in the electrophoretic mobility and abundance of Per (Fig. 1a) similar to those observed for native Per protein in heads from wild-type *Drosophila*² (Fig. 1b; compare lanes 1–3 with lanes 4–6). Daily changes in the apparent molecular mass of native Per are mainly or solely accounted for by differential phosphorylation². Similarly, phosphatase treatment of Per isolated from S2 cells showed that the Dbt-induced changes in mobility were caused by phosphorylation (Fig. 1c). The kinetics of Per hyperphosphorylation is tightly correlated with the formation of stable interactions between Per and Dbt (Fig. 1d). Increasing the levels of *dbt* resulted in steadily faster rates of Per hyperphosphorylation and degradation (data not shown). These results indicate that Dbt interacts stably with Per and functions, at least partly, in a stoichiometric manner, consistent with recent findings in flies¹¹. In the absence of Dbt there is little or no phosphorylation of Per (Fig. 1a; compare lanes 6 and 7).

In the original characterization of *dbt* as a critical kinase functioning in the metabolism of Per, two *dbt* mutants were identified that either lengthen (*dbt^L*) or shorten (*dbt^S*) the periods of clock-controlled rhythmic behaviours⁶, although it is not clear how these alterations might affect activity. Per did not undergo as extensive phosphorylation in the presence of the Dbt-L mutant (Fig. 2a and b, top panels; compare lanes 9–12 with lanes 1–4). Although the

difference was more subtle, hyperphosphorylated isoforms of Per were detected slightly earlier after the induction of Dbt-S than that of wild-type Dbt (Fig. 2a and b; compare lanes 15 and 3). Finally, the progressive multiphosphorylation of Per was essentially abolished with a more severe allele of *dbt* (*dbt^{ar}*) that is associated with arrhythmic locomotor activity in flies¹² (Fig. 2a; compare lanes 5–8 with lanes 1–4). These findings suggest that the effects of the various *dbt* mutants on locomotor activity are a direct consequence of changes in the rate of Per phosphorylation and not secondary issues such as Dbt stability (Fig. 2a and b, bottom panels). Moreover, although it is not known if other kinases contribute to the progressive phosphorylation of Per in flies or in our cultured cells, the results support a central role for Dbt activity in enabling the hyperphosphorylation of Per^{4,6}.

In flies the interaction of Timeless (Tim) with Per stabilizes Per against Dbt-mediated proteolysis^{4,6}. Similarly, in S2 cells the expression of Tim inhibits Dbt-mediated Per hyperphosphorylation and degradation (at 16 h after *dbt* induction there is ~60% more Per in the presence of Tim) (Fig. 2c). Taking these results together, the S2 cell culture system that we employed recapitulates many of the features observed in flies for the Dbt regulation of Per metabolism.

To identify the mechanism underlying the rapid degradation of hyperphosphorylated isoforms of Per, we assayed several well-characterized classes of protease inhibitor. Hyperphosphorylated isoforms of Per accumulated in cells treated with the classic 26S proteasome inhibitors MG115 and MG132 (Supplementary Fig. 1a, and data not shown). In agreement with proteasome-mediated degradation, the Per-associated ubiquitin signal increased approx. fivefold when co-expressed with *dbt* (Supplementary Fig. 1b). This is similar to recent findings showing that in cultured cells mammalian Per proteins are ubiquitinated and degraded by the proteasome^{13,14}. In *Drosophila* the ubiquitin–proteasome pathway is also involved in the light-induced degradation of Tim¹⁵ and the circadian-relevant photoreceptor Cryptochrome (Cry)¹⁶.

SCF (Skp1/Cullin/F-box protein) ubiquitin E3 ligase complexes have been shown to target phosphorylated substrates to the ubiquitin–proteasome pathway^{9,17}. The particular F-box protein present in

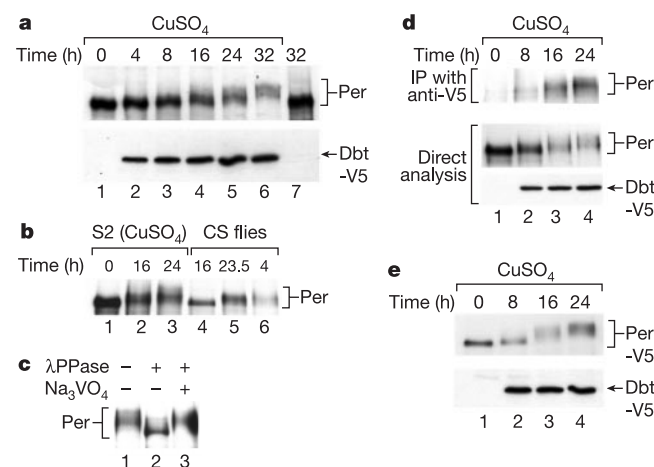


Figure 1 Progressive phosphorylation of Per by Dbt in cultured *Drosophila* cells. S2 cells were co-transfected with pAct-*per* and pMT-*dbt-V5* and extracts were analysed by immunoblotting with antibodies against Per or V5. **a**, Cells were incubated in the presence (lanes 1–6) or absence (lane 7) of CuSO₄ and collected at different times as indicated. Relative levels of Per: lane 1, 100%; lane 2, 78%; lane 3, 69%; lane 4, 60%; lane 5, 51%; lane 6, 39%; lane 7, 113%. IP, immunoprecipitation. **b**, Extracts were prepared from either S2 cells at different times after induction of *dbt-V5* (lanes 1–3) or adult heads isolated from flies collected at different times in a 12 h light–12 h dark cycle (see Methods) (lanes 4–6). **c**, Immune complexes recovered with anti-Per antibodies were incubated in the absence (–) or presence (+) of lambda protein phosphatase (λPPase) and an inhibitor of phosphatase activity (Na₃VO₄). **d**, Extracts were analysed either after immunoprecipitation with anti-V5 antibodies (top panel) or directly (bottom two panels). **e**, Cells were co-transfected with pAct-*per-V5* and pMT-*dbt-V5* and extracts were analysed by immunoblotting with anti-V5 antibodies. The results show that, under our standard transfection conditions, Dbt reached molar amounts 4–6-fold higher than peak values of Per (note that twice as much extract was immunoblotted for the detection of Per than for Dbt-V5).

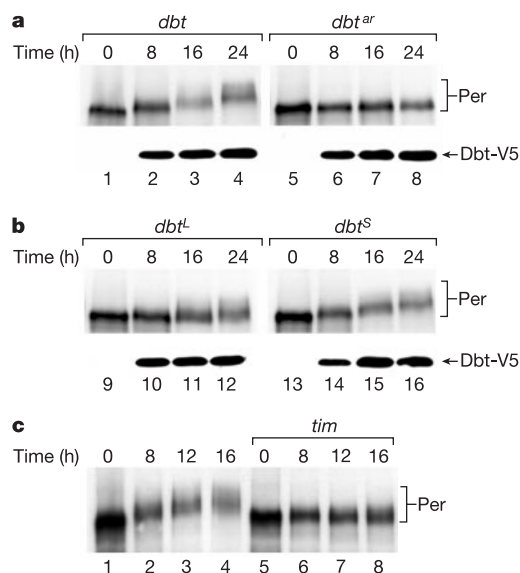


Figure 2 Alterations in Per phosphorylation and degradation by Tim and mutant forms of Dbt. Extracts prepared from S2 cells transfected with pAct-*per* (**a–c**), pMT-*dbt-V5* (**a**, lanes 1–4; **c**), mutant versions of *dbt-V5* under the control of the pMT promoter (**a**, lanes 5–16), and pAct-*tim* (**c**, lanes 5–8) were probed for Per or Dbt-V5 by immunoblotting in the presence of antibodies against Per or V5, respectively. **c**, Half the usual amount of pAct-*per* was used in the transfection.

the SCF complex mainly determines substrate specificity. Among F-box proteins, those that also contain WD (Trp-Asp) repeats have been shown to interact with a wide variety of phosphorylated substrates^{9,17–19}. We therefore sought to determine whether an F-box/WD40 protein might be responsible for the rapid degradation of Dbt-phosphorylated Per. Searches of several databases

revealed the presence of three F-box/WD40 proteins encoded in the *Drosophila melanogaster* genome: *supernumerary limbs* (*slimb*)¹⁰, *archipelago* (*ago*)²⁰ and a predicted gene (GadFly (Genome Annotation Database of *Drosophila*) no. CG9144). Tagged versions of the two characterized F-box/WD40 proteins were expressed in S2 cells and their ability to interact with Per was

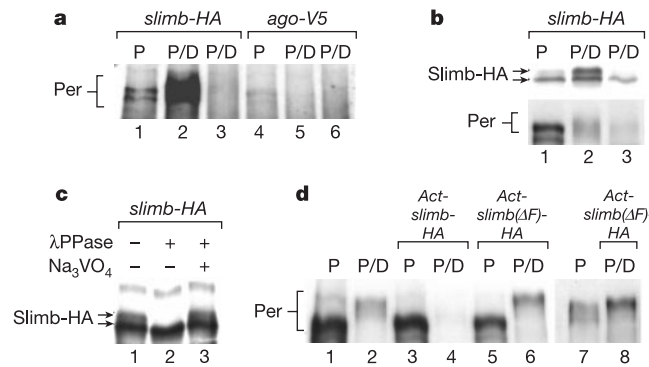


Figure 3 Slimb interacts preferentially with Dbt-phosphorylated Per and mediates its rapid degradation. **a–d**, Cells were transfected with pAct-*per* (P), pMT-*dbt-V5* (D), pMT-*slimb-3HA*, pMT-*ago-V5*, pAct-*slimb-3HA* or pAct-*slimb(ΔF)-3HA* as indicated. Cells were harvested 24 h after the addition of CuSO₄ except in **d**, where they were collected 32 h after induction. Extracts were prepared and either subjected to immunoprecipitation (**a–c**) or analysed directly (**d**). Immunoblots were probed with

antibodies against Per or HA (to detect Slimb-HA), as indicated. Immune complexes were recovered with anti-HA (**a**, lanes 1 and 2; **c**, lanes 1–3), anti-V5 (**a**, lanes 4 and 5), Per (**b**, lanes 1 and 2) or nonspecific (**a**, lanes 3 and 6; **b**, lane 3) antibodies. **c**, Immune complexes were incubated in the absence (–) or presence (+) of lambda protein phosphatase (λPPase) and an inhibitor of phosphatase activity (Na₃VO₄). **b, c**, At least two isoforms of Slimb-HA that differ in mobility were detected (indicated by arrows).

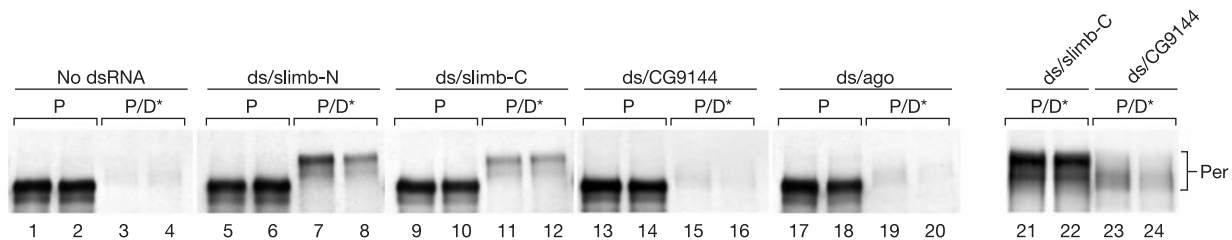


Figure 4 Accumulation of DBT-dependent highly phosphorylated Per by silencing endogenous *slimb* activity. Cells were co-transfected with plasmids containing pAct-*per* (P) or pAct-*dbt-V5* (D*) as indicated. Subsequently, cells were either mock-incubated (lanes 1–4) or incubated with different dsRNAs as indicated (top). Extracts were prepared and probed for Per by western blotting. Two independent experiments were analysed side by side. Autoradiographs exposed for longer periods clearly showed the

presence of more highly phosphorylated variants of Per in the presence of *slimb* dsRNAs compared with no dsRNA or other dsRNAs (for example, lanes 21–24). Endogenous expression of *slimb* in S2 cells and a substantial decrease in its levels by *slimb*-directed RNA interference but not other dsRNAs was confirmed by PCR with reverse transcription (data not shown).

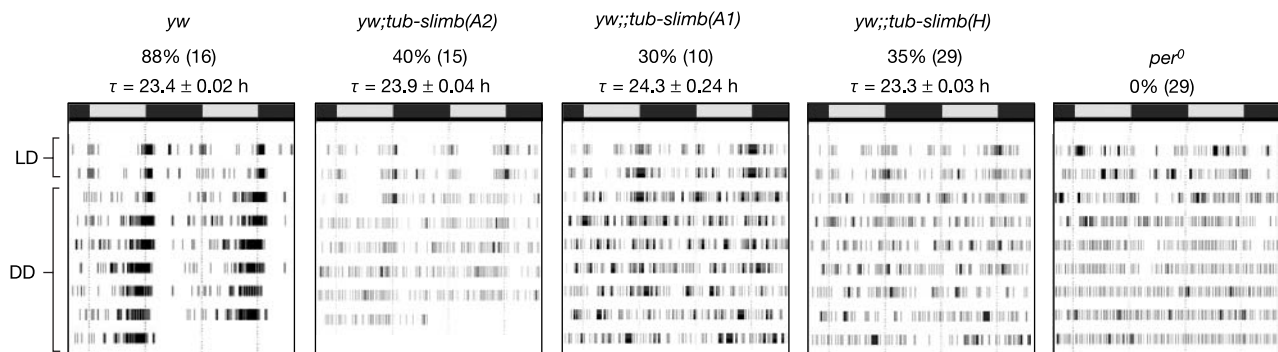


Figure 5 Altered locomotor activity rhythms in *tub-slimb* flies. Shown are representative locomotor activity records of individual flies (actograms) during 3 d of LD followed by 5–7 d of DD, depending on genotype. The 12-h light period and subjective day (grey horizontal bar) and 12-h dark period (black horizontal bar) are indicated. At the top of the panels are genotypes, percentages of rhythmic flies and (in parentheses) the total number

of flies analysed, and average periods of rhythmic flies (τ ; mean $\pm$ s.e.m.). Average 'power' values (measure of the strength of the rhythm) in arbitrary units for each genotype are as follows: *yw*, 38.0 ± 1.8 ; *tub-slimb(A2)*, $12.9 \pm 1.0^{\#}$; *tub-slimb(A1)*, $7.7 \pm 1.1^{\#}$; *tub-slimb(H)*, $13.4 \pm 0.5^{\#}$; *per*⁰¹, $2.5 \pm 0.1^{\#}$ (hash sign denotes significant difference from control *yw* flies, $P < 0.01$ using Student's *t*-test).

determined by immunoprecipitation. Slimb preferentially interacted with Dbt-phosphorylated Per, whereas there was little or no interaction with Ago (Fig. 3a). The preference of Slimb for Dbt-phosphorylated Per cannot be accounted for simply by gross variations in the levels of Per or the different F-box/WD40 proteins (Supplementary Fig. 2). When extracts were incubated with anti-Per antibodies significantly more Slimb co-purified with Dbt-phosphorylated Per (Fig. 3b, top panel), even though higher levels of Per are present in immune complexes recovered from cells not transfected with *dbt* (Fig. 3b, bottom panel; compare lanes 1 and 2).

We noted that Slimb migrates as a doublet (Fig. 3b and c, and Supplementary Fig. 2), which is consistent with recent findings for Slimb²¹ and its vertebrate homologue β -TrCP²². The slower-

migrating isoform of Slimb interacts preferentially with phosphorylated Per (note that a portion of the faster-migrating Slimb co-purified non-specifically) (Fig. 3b). It has been suggested that phosphorylation might underlie the isoforms of Slimb/ β -TrCP²² with different mobilities. Indeed, treatment with phosphatase indicated that the slower-migrating isoform of Slimb is phosphorylated (Fig. 3c). Our results raise the possibility that the phosphorylated state of Slimb/ β -TrCP modulates its ability to interact with target proteins.

High-level expression of *slimb* from a constitutive promoter specifically enhanced the degradation of Dbt-phosphorylated Per (Fig. 3d; compare lanes 4 and 2). The F-box mediates interactions with other core members of the SCF but is not required for the

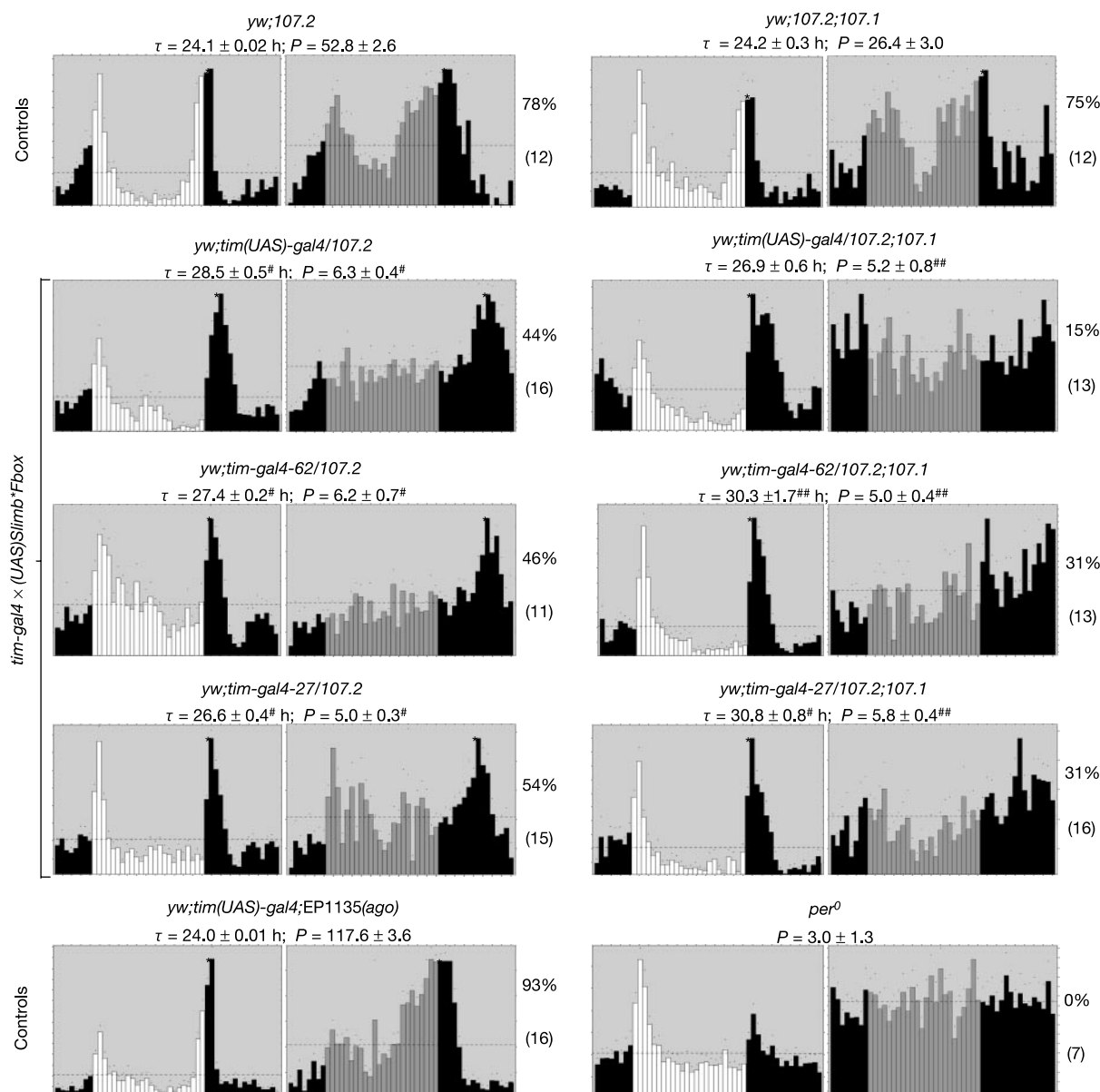


Figure 6 Altered locomotor activity rhythms in *tim-gal4* $\times$ (*UAS*)*Slimb***Fbox* flies. Each panel represents the average activity plots of flies for a given genotype for a consecutive 24-h period during the last day of LD (left, beginning at ZT20) followed by the first day of DD (right, beginning at circadian time 20 h; CT20). At the top of the panels are shown genotypes, periods (τ ; mean $\pm$ s.e.m.) and powers (P ; mean $\pm$ s.e.m.). At the right of panels are shown the percentages of rhythmic flies and (in parentheses) the total numbers of flies analysed. Significant differences in period or power between the

(*UAS*)*Slimb***Fbox* $\times$ *tim-gal4* progeny and their respective controls (107.2 and 107.2;107.1) are indicated as follows: hash sign, $P < 0.01$; double hash sign, $P < 0.05$ using Student's *t*-test. The relatively high proportion of arrhythmic flies in the controls (107.2 and 107.2;107.1) is probably due to leaky expression of *Slimb***Fbox* (data not shown). Vertical bars represent activity recorded in 30-min bins during times when the lights were either on (white bars) or off (black bars, night; grey bars, subjective day during DD). Asterisks indicate peak of evening activity in LD and DD.

specific recognition of phosphorylated substrates^{9,17}. Expression of a Slimb mutant lacking the F-box (Slimb(Δ F)) leads to the accumulation of highly phosphorylated isoforms of Per but only in the presence of Dbt (Fig. 3d, lanes 5–8), strongly suggesting that Slimb(Δ F) is functioning as a dominant-negative inhibitor to block endogenous *slimb* activity. A similar F-box deletion of β -TrCP blocked I κ B α degradation and NF- κ B activation²².

We also used RNA interference²³ to silence any endogenous *slimb* that might be present in our cultured cells (Fig. 4). Incubation of S2 cells with either of two different *slimb* double-stranded RNAs (dsRNAs) efficiently blocked the degradation of phosphorylated Per, whereas dsRNAs directed against the other two *Drosophila* F-box/WD40 proteins had no noticeable effects on Per metabolism. Furthermore, highly phosphorylated isoforms of Per—not normally detected—accumulated in the presence of *slimb* dsRNAs (Fig. 4, lanes 21–24). The addition of *slimb* dsRNAs did not affect the levels of Per in the absence of Dbt, which is consistent with the preferential interaction of Slimb with Dbt-phosphorylated Per (Fig. 3a and b).

To obtain evidence *in vivo* that *slimb* has a role in circadian function we used wild-type flies bearing a transgene in which *slimb* expression was driven from the constitutive *tubulin* promoter (*tub-slimb*). Three independent *tub-slimb* lines (S102.H, S102.A1 and S102.A2) were examined. In standard 12 h light–12 h dark cycles (LD, where zeitgeber time 12 (ZT12) is when the lights are turned off), normal flies exhibit a bimodal distribution of activity centred on the lights-on (morning peak) and lights-off (evening peak) transitions with a robust evening activity rhythm continuing for many days in constant dark conditions (DD). Although all three strains of *tub-slimb* flies entrained (synchronized) to daily light–dark cycles, between 60% and 70% of these flies displayed arrhythmic locomotor activity under constant dark conditions (Fig. 5). Expression of several control genes under the regulation of the *tubulin* promoter did not affect locomotor activity rhythms (data not shown).

In addition, we used the GAL4–UAS binary system (which exploits the yeast GAL4 transcriptional activator and the upstream activating sequence (UAS))²⁴ to drive the production of a mutant form of Slimb lacking the F-box (Slimb*Fbox) in *tim*-expressing cells. We used two independent lines of UAS–Slimb*Fbox (107.1 and 107.2) and the transheterozygote combination (107.1/107.2), and separately crossed each one to three different *tim-gal4* strains (*tim-gal4-27*, *tim-gal4-62* and *tim(UAS)-gal4*)^{25,26}. Flies expressing Slimb*Fbox under the control of the different *gal4* drivers manifested abnormal locomotor activity rhythms in LD (Fig. 6). Notable differences were little or no rise in activity before the beginning of the 12-h dark phase and a delayed downswing in evening activity. More severe differences were observed in DD, in which the majority of the *tim-gal4* $\times$ UAS–Slimb*Fbox flies were arrhythmic or had weak rhythms with long periods, consistent with the delayed evening activity in LD (Fig. 6). These behavioural defects were generally more severe with the stronger *tim(UAS)-gal4* driver^{25,26} and with two copies of the UAS–Slimb*Fbox transgene (Fig. 6). All these mutant phenotypes require *tim-gal4* drivers, because flies carrying the UAS–Slimb*Fbox transgene alone showed normal rhythms (Fig. 6). Furthermore, normal activity rhythms were observed in flies in which the expression of *ago* was regulated by the same *tim-gal4* drivers, indicating that the altered rhythm phenotypes are not due simply to overexpression of any F-box/WD40 protein (Fig. 6). Consistent with the behavioural defects was the western blot analysis, which showed that the temporal regulation of Per abundance and phosphorylation is delayed in flies expressing Slimb*Fbox under the control of the *tim* promoter (data not shown).

We suggest that Dbt-mediated phosphorylation of Per is a key recognition signal that enables Slimb to recruit phosphorylated Per selectively and target it for rapid degradation by the 26S proteasome. Given the highly conserved nature of circadian clocks in

Drosophila and vertebrates it is tempting to speculate that β -TrCP has a similar role in regulating the levels of mammalian Per proteins. □

Methods

Constructs and transfection

The pAct-*per*²⁷, pAct-*tim*²⁷ and pCasper-hs-Ub/HA¹⁶ constructs were described previously. To generate an expression construct in which ubiquitin (Ub) was under the control of an inducible promoter, the Ub/HA octamer from pCasper-hs-Ub/HA was subcloned into the plasmid pMT/V5-His (Invitrogen), generating pMT-Ub/HA. For the *dbt* expression vectors, the coding region of *dbt* was amplified by polymerase chain reaction (PCR) in the presence of the plasmid pGEM-5Zf/*dbt*²⁸ and the relevant fragment was subcloned into either the pAc5.1/V5-His (Invitrogen) or pMT/V5-His vectors to yield pAct-*dbt*-V5, pAct-*dbt* (by introducing an in-frame stop codon upstream of plasmid sequences encoding V5) and pMT-*dbt*-V5. The *dbt* mutant constructs (*dbt*^S, *dbt*^L and *dbt*^{mt}) were generated by site-directed mutagenesis with the QuickChange site-directed mutagenesis kit (Stratagene) and subcloned into the pMT/V5-His vector. Coding sequences for the F-box proteins were amplified by PCR in the presence of the appropriate expressed sequence tag clones (LD08669 for *slimb* and LD21322 for *ago*) and subcloned into the pMT/V5-His vector. The *slimb* open reading frame (ORF) was fused in frame at the carboxy terminus with sequences encoding three tandem HA epitope tags (3HA) followed by a stop codon, yielding pMT-*slimb*-3HA. *slimb*-3HA was also subcloned into the pAct vector to yield pAct-*slimb*-3HA. The plasmid pJM1534 (S. Brill, personal communication) was used in a PCR of the fragment containing the 3HA sequence. In addition, pAct-*slimb*(Δ F)-3HA was generated by site-directed mutagenesis by deleting sequences coding for the region containing amino-acid residues 92–136. The *ago* ORF was fused in-frame with sequences encoding the V5 epitope tag, generating pMT-*ago*-V5. Relevant parts from all final expression vectors were sequenced before use.

The S2 cells and DES expression medium were purchased from Invitrogen, and all procedures were performed in accordance with the manufacturer's instructions. Transfections were performed with Cellfectin (Gibco-BRL) in accordance with the manufacturer's instructions. Induction of genes under the control of the pMT promoter was initiated by adding 500 μ M CuSO₄ to the medium, beginning at 36 h after transfection.

RNA-mediated interference

RNA-mediated interference was applied to S2 cells as described previously²³. After the addition of dsRNA, cells were left to recover for 2 d, transfected with pAct-*per* and pAct-*dbt*-V5 plasmids and incubated for a further 1.5 d before cells were harvested. The dsRNAs used were (numbering of nucleotide sequence begins with start ATG): ds/*slimb*-N, 49–756; ds/*slimb*-C, 404–1102; ds/CG9144, 685–1384; and ds/*ago*, 1–505.

Western blot analysis

Transfected cells were washed with serum-free medium and collected by freezing at the indicated times. For each time point, extracts were prepared and immunoblotted as described previously²⁸. Gels differing in polyacrylamide concentration were used to resolve different proteins (6% for Per and Ago; 12% for Slimb; 8% for experiments involving Ub-HA). Primary antibodies were used at the following dilutions: 12CA5 anti-HA (Boehringer Mannheim) 1:2000, anti-V5 (Invitrogen) 1:6000, and GP73 anti-Per 1:3000 (ref. 29). The intensities of relevant bands on autoradiographs were quantified with a charge-coupled device camera (Alpha Innotech Corporation). Scanned images of autoradiographs were manipulated with Adobe Photoshop.

Immunoprecipitation and treatment with phosphatase

Clarified cell extracts (1 mg total protein in a final volume of 0.5 ml) were subjected to immunoprecipitation with Gammabind Plus (Pharmacia) as described previously²⁸. We added 3 μ l of anti-V5, anti-HA (12CA5) or 5 μ l of anti-Per (GP73) antibodies to the precleared cell extracts. Treatment of immune complexes with phosphatase was performed as described previously²⁸. The phosphatase inhibitor Na₃VO₄ (40 mM final concentration) and/or 200 U of lambda (λ) protein phosphatase (NEB) were added as indicated.

Transgenic flies and locomotor activity rhythms

yw, *per*⁰¹, *tim-gal4-27*, *tim-gal4-62* and *tim*-(UAS)-*gal4* were described previously^{25,26}. The EP line 1135 (Berkeley *Drosophila* Genome Project) was used to overexpress *ago*. To generate the *tub-slimb* construct, cDNA encoding full-length *slimb* was subcloned into the pCS2 vector between *Xho*I and *Xba*I, yielding a *slimb* with six copies of Myc-tag fused at the N-terminus. A *Clal*–*Xba*I fragment containing the Myc-tagged *slimb* coding sequence was then subcloned into a Casper vector containing the 2.6-kilobase tubulin promoter. To generate the Slimb*Fbox construct, sequences encoding an amino-terminal region of Slimb (residues 1–89) and a C-terminal region (residues 144–510) were amplified by PCR with pBluscript-Myc-*slimb* as a template. The two fragments were rejoined and subcloned into the pUAST vector between the *Kpn*I and *Xba*I sites to generate pUAST-Myc-Slimb*Fbox. Slimb*Fbox lacks sequences from residues 90–143. All plasmids were injected into *yw* embryos to generate transformants.

For the analysis of locomotor activity, young adult flies were placed in incubators at 25 °C, exposed to 3 d of 12 h light followed by 12 h dark (12:12LD) and subsequently kept in constant dark conditions (DD) for 5–8 d. Locomotor activity was analysed with the Brandeis Rhythm Package as described previously³⁰. Power is a quantification of the strength of the rhythm during DD^{25,26}. Flies with a power ≥ 5 and a 'width' (number of

period values in 30-min increments above that line) ≥ 2 were considered to be rhythmic^{25,26}.

Received 11 September; accepted 5 November 2002; doi:10.1038/nature01272.

Published online 20 November 2002.

1. Ederly, I. Role of posttranscriptional regulation in circadian clocks: lessons from *Drosophila*. *Chronobiol. Int.* **16**, 377–414 (1999).
2. Ederly, I., Zwiebel, L. J., Dembinska, M. E. & Rosbash, M. Temporal phosphorylation of the *Drosophila* period protein. *Proc. Natl Acad. Sci. USA* **91**, 2260–2264 (1994).
3. Lee, C., Etchegaray, J. P., Cagampang, F. R., Loudon, A. S. & Reppert, S. M. Posttranslational mechanisms regulate the mammalian circadian clock. *Cell* **107**, 855–867 (2001).
4. Kloss, B. *et al.* The *Drosophila* clock gene *double-time* encodes a protein closely related to human casein kinase Ie. *Cell* **94**, 97–107 (1998).
5. Lowrey, P. L. *et al.* Positional syntenic cloning and functional characterization of the mammalian circadian mutation tau. *Science* **288**, 483–492 (2000).
6. Price, J. L. *et al.* *double-time* is a novel *Drosophila* clock gene that regulates PERIOD protein accumulation. *Cell* **94**, 83–95 (1998).
7. Toh, K. L. *et al.* An hPer2 phosphorylation site mutation in familial advanced sleep phase syndrome. *Science* **291**, 1040–1043 (2001).
8. Eide, E. J. & Virshup, D. M. Casein kinase I: another cog in the circadian clockworks. *Chronobiol. Int.* **18**, 389–398 (2001).
9. Kipreos, E. T. & Pagano, M. The F-box protein family. *Genome Biol.* **1**, 30021–30027 (2000).
10. Jiang, J. & Struhl, G. Regulation of the Hedgehog and Wntless signalling pathways by the F-box/WD40-repeat protein Slimb. *Nature* **391**, 493–496 (1998).
11. Kloss, B., Rothenfluh, A., Young, M. W. & Saez, L. Phosphorylation of PERIOD is influenced by cycling physical associations of DOUBLE-TIME, PERIOD, and TIMELESS in the *Drosophila* clock. *Neuron* **30**, 699–706 (2001).
12. Rothenfluh, A., Abodeely, M. & Young, M. W. Short-period mutations of *per* affect a *double-time*-dependent step in the *Drosophila* circadian clock. *Curr. Biol.* **10**, 1399–1402 (2000).
13. Akashi, M., Tsuchiya, Y., Yoshino, T. & Nishida, E. Control of intracellular dynamics of mammalian period proteins by casein kinase Ie (CKIe) and CKIb in cultured cells. *Mol. Cell. Biol.* **22**, 1693–1703 (2002).
14. Yagita, K. *et al.* Nucleocytoplasmic shuttling and mCRY-dependent inhibition of ubiquitylation of the mPER2 clock protein. *EMBO J.* **21**, 1301–1314 (2002).
15. Naidoo, N., Song, W., Hunter-Ensor, M. & Sehgal, A. A role for the proteasome in the light response of the *timeless* clock protein. *Science* **285**, 1737–1741 (1999).
16. Lin, F. J., Song, W., Meyer-Bernstein, E., Naidoo, N. & Sehgal, A. Photoc signaling by cryptochrome in the *Drosophila* circadian system. *Mol. Cell. Biol.* **21**, 7287–7294 (2001).
17. Craig, K. L. & Tyers, M. The F-box: a new motif for ubiquitin dependent proteolysis in cell cycle regulation and signal transduction. *Prog. Biophys. Mol. Biol.* **72**, 299–328 (1999).
18. Koepp, D. M. *et al.* Phosphorylation-dependent ubiquitination of cyclin E by the SCFF^{bw7} ubiquitin ligase. *Science* **294**, 173–177 (2001).
19. Lassot, I. *et al.* ATF4 degradation relies on a phosphorylation-dependent interaction with the SCF^{TRCP} ubiquitin ligase. *Mol. Cell. Biol.* **21**, 2192–2202 (2001).
20. Moberg, K. H., Bell, D. W., Wahner, D. C., Haber, D. A. & Hariharan, I. K. Archipelago regulates cyclin E levels in *Drosophila* and is mutated in human cancer cell lines. *Nature* **413**, 311–316 (2001).
21. Bocca, S. N., Muzzopappa, M., Silberstein, S. & Wappner, P. Occurrence of a putative SCF ubiquitin ligase complex in *Drosophila*. *Biochem. Biophys. Res. Commun.* **286**, 357–364 (2001).
22. Spencer, E., Jiang, J. & Chen, Z. J. Signal-induced ubiquitination of IκBα by the F-box protein Slimb/β-TrCP. *Genes Dev.* **13**, 284–294 (1999).
23. Worby, C. A., Simonson-Leff, N. & Dixon, J. E. RNA interference of gene expression (RNAi) in cultured *Drosophila* cells. *Sci. STKE* **2001**, PL1 (2001).
24. Brand, A. H. & Perrimon, N. Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* **118**, 401–415 (1993).
25. Kaneko, M., Park, J. H., Cheng, Y., Hardin, P. E. & Hall, J. C. Disruption of synaptic transmission or clock-gene-product oscillations in circadian pacemaker cells of *Drosophila* cause abnormal behavioral rhythms. *J. Neurobiol.* **43**, 207–233 (2000).
26. Blau, J. & Young, M. W. Cycling *vriille* expression is required for a functional *Drosophila* clock. *Cell* **99**, 661–671 (1999).
27. Ceriani, M. F. *et al.* Light-dependent sequestration of TIMELESS by CRYPTOCHROME. *Science* **285**, 553–556 (1999).
28. Bae, K., Lee, C., Hardin, P. E. & Ederly, I. dCLOCK is present in limiting amounts and likely mediates daily interactions between the dCLOCK-CYC transcription factor and the PER–TIM complex. *J. Neurosci.* **20**, 1746–1753 (2000).
29. Sidote, D., Majercak, J., Parikh, V. & Ederly, I. Differential effects of light and heat on the *Drosophila* circadian clock proteins PER and TIM. *Mol. Cell. Biol.* **18**, 2004–2013 (1998).
30. Kim, E. Y. *et al.* *Drosophila* CLOCK protein is under posttranscriptional control and influences light-induced activity. *Neuron* **34**, 69–81 (2002).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank S. Kay for the pAct-*per* and pAct-*tim* constructs, A. Sehgal for the pCasper-hs-Ub/HA construct, J. Hall for the *tim-gal4*-27 and *tim-gal4*-62 lines, and M. Young for the *tim*-(UAS)-*gal4* line. The work was supported by the Searle Scholar Program and the UTSW Endowed Scholar Program to J.J., and by a grant from the National Institutes of Health to I.E.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.E. (e-mail: edery@cabm.rutgers.edu).

HIV-1 evades antibody-mediated neutralization through conformational masking of receptor-binding sites

Peter D. Kwong^{†‡}, Michael L. Doyle^{‡§}, David J. Casper[‡], Claudia Cicala^{||}, Stephanie A. Leavitt[¶], Shahzad Majeed^{†‡}, Tavis D. Steenbeke^{||}, Miro Venturi^{*}, Irwin Chaiken[#], Michael Fung[☆], Hermann Katinger^{**}, Paul W. I. H. Parren^{†‡}, James Robinson^{‡‡}, Donald Van Ryk[#], Liping Wang^{§§}, Dennis R. Burton^{†‡}, Ernesto Freire[¶], Richard Wyatt^{§§}, Joseph Sodroski^{§|||}, Wayne A. Hendrickson^{†¶} & James Arthos^{||}

^{*} Vaccine Research Center, and ^{||} National Institute of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, Maryland 20892, USA

[†] Department of Biochemistry and Molecular Biophysics, and [¶] Howard Hughes Medical Institute, Columbia University, New York, New York 10032, USA

[‡] Department of Structural Biology, GlaxoSmithKline Pharmaceuticals, King of Prussia, Pennsylvania 19406, USA

[¶] Department of Biology, Johns Hopkins University, Baltimore, Maryland 21218, USA

[#] Department of Medicine, University of Pennsylvania, Philadelphia, Pennsylvania 19104, USA

[☆] Cell Biology, Tanox, Houston, Texas 77025, USA

^{**} Institute for Applied Microbiology, University of Agriculture and Forestry, A-1190 Vienna, Austria

^{††} Departments of Immunology and Molecular Biology, Scripps Research Institute, La Jolla, California 92037, USA

^{‡‡} Department of Pediatrics, Tulane University Medical Center, New Orleans, Louisiana 70112, USA

^{§§} Department of Cancer Immunology and AIDS, Dana-Farber Cancer Institute, and Department of Pathology, Division of AIDS, Harvard Medical School, and ^{|||} Department of Immunology and Infectious Diseases, Harvard School of Public Health, Boston, Massachusetts 02115, USA

The ability of human immunodeficiency virus (HIV-1) to persist and cause AIDS is dependent on its avoidance of antibody-mediated neutralization. The virus elicits abundant, envelope-directed antibodies that have little neutralization capacity¹. This lack of neutralization is paradoxical, given the functional conservation and exposure of receptor-binding sites on the gp120 envelope glycoprotein, which are larger than the typical antibody footprint² and should therefore be accessible for antibody binding. Because gp120–receptor interactions involve conformational reorganization³, we measured the entropies of binding for 20 gp120-reactive antibodies. Here we show that recognition by receptor-binding-site antibodies induces conformational change. Correlation with neutralization potency and analysis of receptor–antibody thermodynamic cycles suggested a receptor-binding-site ‘conformational masking’ mechanism of neutralization escape. To understand how such an escape mechanism would be compatible with virus–receptor interactions, we tested a soluble dodecameric receptor molecule and found that it neutralized primary HIV-1 isolates with great potency, showing that simultaneous binding of viral envelope glycoproteins by multiple receptors creates sufficient avidity to compensate for such masking. Because this solution is available for cell-surface receptors but not for most antibodies, conformational masking enables HIV-1 to maintain receptor binding and simultaneously to resist neutralization.

The gp120 envelope glycoprotein binds sequentially to the cellular receptors CD4 and a member of the chemokine receptor

[§] Present address: Biopharmaceuticals Department H13-07, Bristol-Myers Squibb PRI, Pharmaceutical Research Institute, P.O. Box 4000, Princeton, New Jersey 08543-4000, USA.

Table 1 Thermodynamic parameters of binding to HIV-1 gp120

Ligand	gp120	ΔG (kcal mol ⁻¹)	ΔH (kcal mol ⁻¹)	$-T\Delta S$ (kcal mol ⁻¹)
CD4	YU2	-10.4 ± 0.1	-54.6 ± 1.9	44.2 ± 1.9
	HXBc2 core	-10.2 ± 0.1	-48.4 ± 2.5	38.2 ± 2.5
17b	YU2	-10.6 ± 0.9	-44.9 ± 4.0	34.3 ± 4.0
	YU2 (30 °C)	-10.8 ± 0.5	-32.6 ± 3.0	21.8 ± 3.0
48d	YU2 (25 °C)	-10.8 ± 0.2	-27.4 ± 3.0	16.6 ± 3.0
	YU2	-10.3 ± 0.2	-39.0 ± 3.9	28.7 ± 3.9
F105	HXBc2 core	-11.4 ± 0.2	-30.0 ± 3.0	18.6 ± 3.0
b12	YU2	-11.1 ± 0.2	-17.0 ± 1.7	5.9 ± 1.7
	HXBc2 core	-12.3 ± 0.3	-18.0 ± 1.8	5.7 ± 1.8
2G12	HXBc2 core	-7.8 ± 0.2	-6.2 ± 0.6	-1.6 ± 0.6
C11	YU2 (+b12)	-12.5 ± 0.3	-13.5 ± 1.6	1.0 ± 1.6
	YU2 (+CD4 + 17b)	-12.2 ± 0.3	-16.0 ± 1.6	3.8 ± 1.6

All measurements were made at 37 °C except where specified.

family (CXCR4 or CCR5) to initiate virus cell entry⁴⁻⁶. We used isothermal titration calorimetry to measure the thermodynamics of binding of monomeric gp120 to CD4 and six different antibodies^{3,7-12} (Table 1). Measurements were made with both core HXBc2 and full-length YU2 gp120. Free energies (ΔG) of interaction were similar, ranging from -10.2 to -12.5 kcal mol⁻¹ (1 kcal = 4.18 kJ) with an average deviation from the mean of 0.9 (except for 2G12, which showed lower than expected binding, probably as a result of altered carbohydrate on the *Drosophila*-expressed core^{11,13}). In contrast, enthalpy (ΔH) and entropy ($-T\Delta S$) were widely divergent, each covering an almost 50 kcal mol⁻¹ range with an average deviation from the mean of 15 kcal mol⁻¹. Nevertheless, a plot of ΔH against $-T\Delta S$ in Table 1 indicates an almost perfect enthalpy/entropy compensation with a slope of -0.98 and a correlation coefficient of 0.997. Core and full-length gp120 exhibited comparable thermodynamic changes upon binding ligands (see CD4 and b12; Table 1). Thermodynamic parameters were similar for the binding of C11 (discontinuous epitope at the gp41-interactive region on the inner domain) to preformed gp120-b12 and gp120-CD4-17b complexes. The results demonstrate that not only CD4, but also antibody ligands, can induce substantial enthalpic and entropic changes in gp120, and moreover that these changes are a direct reflection of the intrinsic effect of each ligand on the gp120 core, not related to details of isolate, the presence of variable loops, or binding at the gp41-interactive region.

To study a larger range of ligands, a novel 'deglycosylation assay' was devised on the basis of the sensitivity of degradative enzymes to the degree of folding of the substrate, with extracted entropies from

kinetic measurements based on empirical correlations observed with a subset of the antibodies (details and theoretical considerations are described in Methods and Supplementary Information). By using this deglycosylation assay to complement and extend calorimetry results, the entropy of binding to monomeric gp120 was determined for 20 monoclonal antibodies (Table 2). The antibodies tested reacted with a broad spectrum of gp120 epitopes. Variation in both gp120 strain and construct was used to ensure generality in the results.

Correlations between thermodynamic parameters and epitope identity were analysed. No correlations were observed with free energy (ΔG) of binding. In contrast, a striking correlation between unusual entropy and receptor-binding-site epitope was observed. Receptor-binding-site antibodies exhibited large negative changes in entropy that amounted to an average contribution of 23.1 kcal mol⁻¹ to the free energy of gp120-antibody binding, whereas non-receptor-binding-site antibodies displayed an average contribution of only 0.4 kcal mol⁻¹ (Table 2). For comparison, the typical value for antibody-protein binding is 7 ± 3 kcal mol⁻¹ (ref. 14), and variable-loop-reactive antibodies and amino/carboxy-terminus-reactive antibodies had average entropies of -2.8 ± 3.3 and 5.2 ± 7.1 kcal mol⁻¹, respectively.

How can these unusual entropies be explained in terms of the gp120 structure (Fig. 1)? Large negative entropies are indicative of surface burial or protein folding. Estimates based on the CD4-gp120 interaction³ suggest that for gp120-reactive antibodies with unusual entropies, 4,000-7,000 Å² of surface is buried or 40-70 amino acids are folded. To gain a more precise measure of the antibody-induced conformational change, isothermal titration calorimetry experiments of the YU2 gp120-17b interaction were performed at different temperatures (Table 1). These yielded a change in heat capacity (ΔC_p) of -1.5 kcal K⁻¹ mol⁻¹, which is consistent with the burial of $\sim 9,800$ Å², about 10-fold larger than the 1,000 Å² calculated from the crystallographic structure¹⁵ and consistent with the ordering of 65 amino-acid residues¹⁶. Thermodynamic analysis of monomeric gp120 indicates that it is highly flexible, with the bridging sheet unfolded and a mobile interface between the inner and outer domains³. The receptor-binding sites are sandwiched at the interface of the inner and outer domains, between surfaces occluded by oligomerization and glycosylation, respectively. To achieve the necessary surface area for high-affinity binding, receptor-binding-site antibodies must bridge domains,

Table 2 Entropy of antibody binding to HIV-1 gp120

Antibody epitope competition group	Neutralization phenotype*	Monoclonal antibody	− TΔS (kcal mol ^{−1})†	gp120 construct
Receptor-binding-site-reactive	Broadly reactive			
Chemokine-receptor-binding site (CD4i)	Non-potent	17b	31.5 ± 2.8	YU2
	Non-potent	48d	28.7 ± 3.9	YU2
CD4-binding site (CD4BS)	Non-potent	1.5e	27.3 ± 7.7	YU2 core + V3
	Non-potent	b6	27.7 ± 11.9	YU2 core + V3
	Potent	b12	5.7 ± 1.8	HXBc2 core
	Non-potent	F105	18.6 ± 3.0	HXBc2 core
	Non-potent	F91	22.8 ± 6.0	JRFL core + V3
Variable-loop-reactive	Strain-specific			
V1/2 loop	Specific	L17	−7.8 ± 8.1	HXBc2
	Specific	L78	−2.7 ± 6.2	HXBc2
V3 loop	Specific	19b	−6.5 ± 3.7	JRFL core + V3
	Specific	39F	−2.6 ± 6.1	YU2 core + V3
	Specific	Ag1211	−0.1 ± 3.9	JRFL core + V3
	Specific	D0142	0.5 ± 9.0	YU2 core + V3
V3 loop/outer domain	Specific	G3-299	−0.1 ± 7.8	YU2 core + V3
Carbohydrate-reactive	Broadly reactive			
Outer domain	Potent	2G12	−1.6 ± 0.6	HXBc2 core
N/C-termini-reactive	Non-neutralizing			
N terminus, discontinuous	Non-neutralizing	211/c	10.1 ± 4.8	HXBc2
	Non-neutralizing	A32	13.4 ± 5.2	HXBc2
	Non-neutralizing	L100	−1.5 ± 3.9	HXBc2
N terminus, linear	Non-neutralizing	P35	−2.6 ± 3.5	HXBc2
N and C termini	Non-neutralizing	C11	6.7 ± 4.3	HXBc2

* 'Broadly reactive' antibodies bind to diverse strains of HIV-1. These may be either 'potent', able to neutralize primary isolates, or 'non-potent', weakly neutralizing even laboratory-adapted isolates. 'Strain-specific' antibodies bind in a strain-restricted manner, whereas 'non-neutralizing' antibodies recognize only monomeric gp120.

† Normalized per mole of gp120; 37 °C isothermal titration calorimetry reported for 17b, 2G12, 48d, b12 and F105, and the deglycosylation assay for all others; multiple measurements are averaged.

necessitating their ordering and incurring an entropic penalty. Thus, the size of the antibody footprint, constraints on antibody orientation imposed by glycosylation and oligomeric occlusion, and the domain organization of gp120 in conjunction with its intrinsic flexibility all combine to impose an entropic/conformational

barrier on antibodies directed against receptor-binding regions.

Correlations between entropy and neutralization potency were also analysed. The entropy measurements described here were made within the context of the flexible monomeric gp120, where unusual unfavourable entropies are compensated for by large enthalpic gains involving substantial conformational change. Neutralization, by contrast, occurs within the context of the functional oligomeric spike, where quaternary constraints are likely to restrict conformational change. Numerous studies have shown that antibody binding to oligomeric, but not monomeric, gp120 is correlated with neutralization (for example, see ref. 17). Accordingly, correlation with neutralization provides a method of examining the biological consequences of the intrinsic core changes associated with ligand binding that we have revealed here. With broadly reactive antibodies we observed a remarkable correlation between entropy and neutralization potency. Non-potent but broadly reactive antibodies all showed unusual entropies (a mean entropy contribution to free energy of $26.1 \text{ kcal mol}^{-1}$; Table 2), indicating a thermodynamic mechanism for neutralization resistance, with quaternary constraints restricting the conformational reorganization required for the binding of antibodies with unusual entropy. The high compensating values for entropy and enthalpy allow ample options for the manipulation of free energy, and thus the affinity, of receptor-binding-site ligands in the context of the envelope glycoprotein trimer, with structural analysis indicating that the determinants of neutralization resistance reside in variable loops involved in quaternary interactions¹⁵.

If this quaternary/entropic explanation were a primary mechanism by which HIV avoids neutralization, then we would expect broadly reactive antibodies that potently neutralize primary isolates to bind with a small change in entropy. Of the many anti-gp120

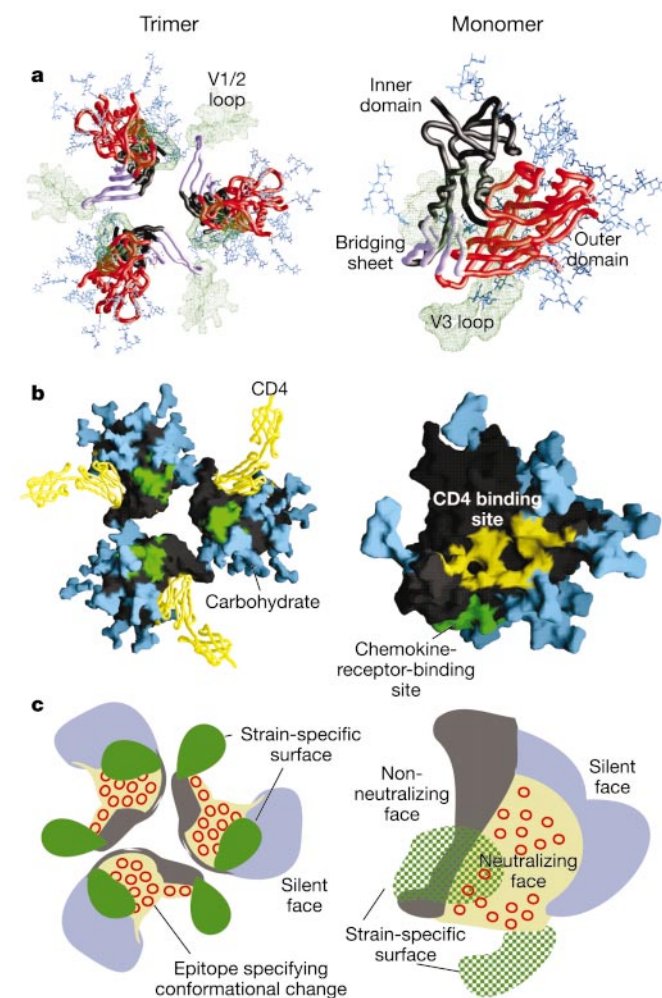


Figure 1 HIV-1 gp120: structure, receptor binding, antigenicity and entropy of antibody binding. The envelope glycoproteins gp120 and gp41 assemble into a trimeric heterodimer on the surface of HIV-1. The left panels show gp120 as positioned by the optimization of quantifiable surface parameters into the orientation that it most probably assumes in the viral spike²⁷. The view shown here corresponds to that seen from the target cell membrane. The right panels show monomeric gp120. Its orientation corresponds to the rightmost protomer of the trimer rotated 90° about a horizontal axis such that the target cell membrane is positioned below, the viral membrane above. **a**, Atomic structure of gp120. The gp120 core is displayed as a Cα worm with inner domain coloured black, the outer domain red, and the bridging sheet violet. The approximate positions of the V1/2 and V3 variable loops are shown in green, and the protein-proximal (mannose)₃ glycan cores as modelled previously²⁴ are shown in blue. **b**, Functional surface of gp120. The molecular surface of HXBc2 gp120 is coloured by functionality: yellow, surface within 3.5 Å of CD4; green, surface associated with residues identified by mutagenesis¹² as being part of the chemokine-receptor-binding site; blue, N-linked carbohydrate; black, remaining gp120 protein surface. A Cα worm (yellow) of the N-terminal domains of CD4 is shown for reference as positioned by the atomic structure of the HXBc2 gp120-core-CD4 complex². **c**, Antigenic structure. The antigenic structure is shown corresponding to epitopes bound by non-neutralizing antibodies²⁸ (black), neutralizing antibodies that are broadly reactive^{7-9,11,12} (light green) and neutralizing antibodies that are strain specific^{29,30} (square pattern in forest green). The highly glycosylated gp120 surface that is rarely recognized by antibodies is coloured blue. Red circles designate those epitopes that display unusual entropy indicative of conformational change. **a** and **b** were drawn with GRASP (authors A. Nicholls, K.A. Sharp and B. Honig).

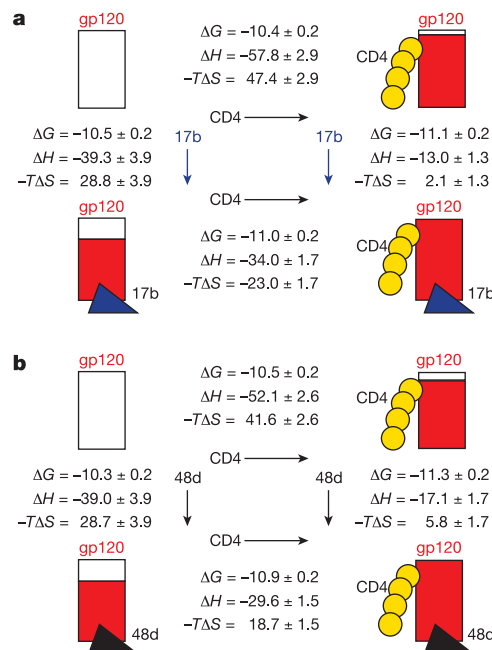


Figure 2 Thermodynamic cycles of 17b and 48d with CD4. Isothermal titration calorimetry was used to measure the thermodynamic parameters at 37 °C for the ternary interactions of YU2 gp120, CD4, and the CD4i antibodies 17b and 48d. **a**, Thermodynamic cycle for 17b. The thermodynamic parameters are shown for CD4 additions (horizontal reactions) and 17b additions (vertical reactions). **b**, Thermodynamic cycle for 48d. Because ΔG , ΔH and $T\Delta S$ are all state functions, these cycles permit the deconvolution of the conformational changes induced in gp120 by the binding of each ligand is depicted in red. (Results are shown for individual experiments. Variations reflect experimental error between experiments.)

antibodies analysed for their ability to neutralize HIV-1, two—b12 and 2G12—are unusually potent, effective even in inhibiting relatively resistant primary HIV-1 isolates. Antibody b12 recognizes the CD4-binding site⁹, whereas 2G12 recognizes a carbohydrate-dependent epitope located on the gp120 outer domain¹¹. Both antibodies exhibited small binding entropies to gp120 (Table 2). This low entropy of binding is the first biochemical property that clearly distinguishes b12 from other CD4BS antibodies.

To verify that non-potent receptor-binding site antibodies induce conformational changes in the viral spike, we investigated the complete thermodynamic cycles of ternary complex formation for gp120 binding to CD4 and to the CD4i antibodies 17b and 48d^{8,12} (Fig. 2). Initial binding by CD4 induced conformational changes in gp120 that reduced the entropy of 17b and 48d binding by 26.7 and 22.9 kcal mol⁻¹, respectively; CD4 binding induced a conformation to which 17b and 48d bound with little entropic change. Initial binding by 17b or 48d induced conformational changes in gp120 that reduced the entropy of subsequent CD4 binding by 24.4 and 22.9 kcal mol⁻¹ for 17b and 48d, respectively. This initial CD4i-antibody binding induced an entropic change that was roughly two-thirds the magnitude of the CD4-induced change (about 65 and 100 residues become structured in each case).

The 17b and 48d thermodynamic cycles displayed unusual, but remarkably similar, thermodynamics. To test whether this similarity was related to a similarity in the antibodies themselves, we determined their amino acid sequences. The sequence of 48d was shorter than that of 17b in both the heavy chain (nine amino acids shorter) and the light chain (three amino acids shorter) for the critical third complementary-determining regions, which make up two-thirds of the 17b-gp120 binding interface¹⁵; indeed, none of the 16 residues of 17b that contact gp120 are conserved in 48d (Supplementary Information). The results indicate that the unusual thermodynamics are a property of the gp120 epitope and that the observed induction of CD4i antibody binding by CD4 is due not merely to the exposure of a gp120 loop or the removal of a steric block but to a CD4-induced conformational rearrangement of gp120 that involves the structuring of a large number of core residues. They demonstrate that CD4i antibodies require conformational changes in gp120 to bind, and that these changes can be fully induced by CD4.

Numerous studies have documented the neutralization synergy between CD4 and CD4i antibodies. Sub-neutralizing concentrations of CD4 potentiate the ability of CD4i antibodies to neutralize primary isolates¹⁸; the presence of soluble CD4 allows 17b to block the formation of primary isolate syncytia¹⁹; and truncated versions of 17b, but not the full-length antibody, neutralize primary isolates (D.R.B., unpublished observations). In this last instance, temperature-shift experiments show that neutralization occurs at the sterically restricted gap between the CD4-attached virus and the target cell membrane, after cell-surface CD4 has induced the requisite conformational change in the viral spike (J.S., unpublished observations). These studies all support the

notion that restrictions in conformational change in the context of the virion oligomer prevent antibody binding and neutralization at sites of receptor binding.

The conformational masking mechanism of neutralization escape that we propose here should affect all ligands that bind to the receptor binding sites. This includes CD4, providing a rationale for the observation that most primary isolates bind soluble CD4 with reduced affinity²⁰. So how does HIV bind to cell-surface CD4 to initiate entry? We suggest that HIV-1 takes advantage of the ability of cell-surface CD4 to bind multiple sites on envelope trimers simultaneously. We tested a dodecameric CD4 molecule (D1D2-Igαtp), created by placing the membrane-distal two domains of CD4 into an IgG-IgA tailpiece context. Analysis of gp120-D1D2-Igαtp binding kinetics and stoichiometry suggested a potential for a highly avid interaction²¹. Although several oligomeric forms of CD4 had been created previously, dimeric and tetrameric forms were unable to bind simultaneously to three CD4-binding sites within a functional viral spike²², and a multimeric CD4-IgM chimaera was never tested with primary isolates²³. We compared the neutralization ability of monomeric CD4 (sCD4) with D1D2-Igαtp to remarkable effect (Table 3); 90%-inhibitory concentrations (IC₉₀) for all primary HIV-1 isolates were achieved at less than 3 nM D1D2-Igαtp.

The properties of the HIV-1 envelope glycoproteins that simultaneously allow the retention of function and the evasion of the humoral immune response are gradually becoming apparent. Antigenic and structural analyses of the most exposed component of the trimeric envelope glycoprotein complex, gp120, reveal four surfaces: a non-neutralizing face, a variable surface, a neutralizing face and an immunologically 'silent' face²⁴. The non-neutralizing face, the variable surface and the 'silent' face are protected from the immune system by occlusion on the oligomer, by mutational variation and by carbohydrate masking, respectively. Our results suggest that the potentially vulnerable receptor-binding sites on gp120 are protected by a novel type of camouflage: conformational or entropic masking. The consequence of this conformational mask is that antibodies against the receptor-binding regions incur an energetic handicap to binding not faced by other anti-gp120 antibodies. This energetic barrier might influence the efficiency with which receptor-binding-site antibodies are generated as well as the neutralization efficacy of these antibodies. Additional tertiary/quaternary structural features of the envelope glycoprotein spikes on primary HIV-1 isolates are thought to provide further constraints on antibody interaction with receptor-binding sites, while allowing sufficient affinity/avidity for receptors to support infection. An understanding of the properties of the HIV-1 envelope that permit neutralization resistance will guide attempts to create vaccines as well as therapeutics that target receptor binding—as we have done with the extraordinarily potent D1D2-Igαtp. □

Methods

Isothermal titration calorimetry

Isothermal titration calorimetry (ITC) measurements were performed at 37 °C with a MicroCal MCS or VP instrument (Amherst, MA) and data were analysed with Origin software (MicroCal, Inc.). Solution conditions were 10 mM sodium phosphate, 200 mM NaCl, 0.2 M EDTA, pH 7.4. Antibodies or CD4 were injected from a stirring syringe into a calorimeter cell containing gp120. Concentrations of gp120 proteins were initially determined by absorbance at 280 nm. Concentrations of antibodies were initially approximated using the Bradford assay to design titration experiments, and were subsequently defined during data analysis by using CD4 as a reference. The concentration of CD4 was determined by absorbance to an accuracy of ±5% as described previously²⁵ and then used subsequently to determine the active concentration of gp120. The binding thermodynamic parameters ΔH and $-T\Delta S$ were therefore measured in units of kcal mol⁻¹ of active gp120.

The sizable quantities of material required for isothermal titration calorimetry made it difficult to analyse a panel of antibodies. Several other methods including tryptophan fluorescence, deuterium exchange with mass spectroscopy, hydrophobic dye binding and enzymatic modification were attempted to develop an assay that required less material. We exploited the observation that the rate of deglycosylation is dependent on the intrinsic flexibility of the substrate, with denatured substrates deglycosylating much more rapidly

Table 3 Neutralization of primary HIV-1 isolates by D1D2Ig-αtp and sCD4

Isolate	IC ₉₀ , D1D2Ig-αtp		IC ₉₀ , sCD4	
	(nM)	(μg ml ⁻¹)	nM	(μg ml ⁻¹)
pi102	0.66 ± 0.42	0.53 ± 0.33	>1,200	>50
pi104	2.72 ± 0.63	2.18 ± 0.46	>1,200	>50
pi202	1.16 ± 0.57	0.93 ± 0.44	>1,200	>50
pi204	1.72 ± 0.66	1.37 ± 0.51	>1,200	>50
Bal	0.12 ± 0.02	0.10 ± 0.02	280 ± 50	11 ± 2
JRFL	0.14 ± 0.02	0.11 ± 0.01	>1,200	>50

Minimally passaged primary isolates of HIV-1 (obtained from individuals within 12 months after seroconversion) and molecularly cloned Bal and JRFL were added to activated peripheral blood mononuclear cells along with either D1D2-Igαtp or sCD4. Reverse transcriptase present in the cell supernatants was measured each day. The reverse transcriptase activity on the day before peak replication in the untreated control culture was used to determine the percentage inhibition in the treated cultures. The concentration (IC₉₀) of D1D2Ig-αtp and sCD4 that inhibited 90% of virus replication is reported.

than well-folded substrates (see Supplementary Information). This permitted the development of a gel-based assay that used susceptibility to deglycosylation as an indicator of the entropic state of gp120, either alone or in complex with an antibody.

Deglycosylation

Deglycosylations were performed with gp120 produced in *Drosophila* Schneider 2 lines. Carbohydrate analysis of this protein showed that about 70% of the attached sugars were sensitive to digestion by endoglycosidase H (Endo H), the rest being a smaller (mannose)₃ variety. The gp120 proteins included core + V3 loop from both YU2 and JRFL, and full-length HXBc2. Mass-spectroscopy analysis of full-length HXBc2 before and after digestion by Endo H showed relative molecular masses of 84,200 ± 2,600 and 64,000 ± 1,600, respectively.

Deglycosylation reactions were performed in 350 mM NaCl under a variety of different conditions with pH varying between 5.5 and 6.5, reaction times between 2 and 20 h, and gp120 concentrations between 0.1 and 1.0 mg ml⁻¹. Reactions were generally performed at 37 °C; reducing the temperature to 20 °C altered the slope of the normalization but not the degree of fit. Ligands were added to a minimum ligand-binding-site:gp120 stoichiometry of 1.5:1 to ensure the complete binding of all gp120. Ligands were CD4 (domains 1 and 2), purified monoclonal antibodies, or antigen-binding portions of antibodies.

Results from five separate deglycosylation experiments are reported here, two with full-length HXBc2, two with YU2 (core + V3 loop) and one with JRFL (core + V3 loop). In each of these experiments, 5–10 different ligands were tested, of which at least half had entropies determined by isothermal titration calorimetry and could be used for calibration. The ligand make-up and relative rates of deglycosylation for each experiment are given in the Supplementary Information.

Sequencing

Sequencing of 17b and 48d antibodies was accomplished with primers that encoded the N termini of the variable domains and the conserved constant-region hinge sequences. Sequences and methodology are detailed in the Supplementary Information.

Neutralization

D1D2-IgGtp was made as described previously²¹. Neutralization assays were performed on freshly isolated activated peripheral-blood mononuclear cells as described²⁶. Activated cells (200,000) were inoculated with 100 tissue-culture-infectious-dose-50 units of each isolate, with the simultaneous addition of sCD4, D1D2-IgGtp or PBS. Virus was washed from the cells after 24 h. Cultures were maintained for 15 days in the presence of interleukin-2-containing medium. Reverse-transcriptase activity was measured on a day before peak replication in the control supernatants. All cultures were performed in quintuplicate.

Received 28 May; accepted 23 September 2002; doi:10.1038/nature01188.

- Weiss, R. A. *et al.* Neutralization of human T-lymphotropic virus type III by sera of AIDS and AIDS-risk patients. *Nature* **316**, 69–72 (1985).
- Kwong, P. D. *et al.* Structure of an HIV gp120 envelope glycoprotein in complex with the CD4 receptor and a neutralizing human antibody. *Nature* **393**, 648–659 (1998).
- Myska, D. G. *et al.* Energetics of the HIV gp120-CD4 binding reaction. *Proc. Natl Acad. Sci. USA* **97**, 9026–9031 (2000).
- Dalglish, A. G. *et al.* The CD4 (T4) antigen is an essential component of the receptor for the AIDS retrovirus. *Nature* **312**, 763–767 (1984).
- Feng, F., Broder, C. C., Kennedy, P. E. & Berger, E. A. HIV-1 entry co-factor: functional cDNA cloning of a seven-transmembrane, G protein-coupled receptor. *Science* **272**, 872–877 (1996).
- Klatzmán, D. *et al.* T-lymphocyte T4 molecule behaves as the receptor for human retrovirus LAV. *Nature* **312**, 767–768 (1984).
- Posner, M. R. *et al.* An IgG human monoclonal antibody that reacts with HIV-1/GP120, inhibits virus binding to cells, and neutralizes infection. *J. Immunol.* **146**, 4325–4332 (1991).
- Thali, M. *et al.* Characterization of conserved human immunodeficiency virus type 1 (HIV-1) gp120 neutralization epitopes exposed upon gp120-CD4 binding. *J. Virol.* **67**, 3978–3988 (1993).
- Burton, D. R. *et al.* Efficient neutralization of primary isolates of HIV-1 by a recombinant human monoclonal antibody. *Science* **266**, 1024–1027 (1994).
- Moore, J. P. & Sodroski, J. Antibody cross-competition analysis of the human immunodeficiency virus type 1 exterior envelope glycoprotein. *J. Virol.* **70**, 1863–1872 (1996).
- Trkola, A. *et al.* Human monoclonal antibody 2G12 defines a distinctive neutralization epitope on the gp120 glycoprotein of human immunodeficiency virus type 1. *J. Virol.* **70**, 1100–1108 (1996).
- Rizzuto, C. D. *et al.* A conserved human immunodeficiency virus gp120 glycoprotein structure involved in chemokine receptor binding. *Science* **280**, 1949–1953 (1998).
- Binley, J. M. *et al.* Analysis of the interaction of antibodies with a conserved, enzymatically deglycosylated core of the HIV-1 gp120 envelope glycoprotein. *AIDS Res. Hum. Retroviruses* **14**, 191–198 (1998).
- Stites, W. E. Protein-protein interactions: interface structure, binding thermodynamics, and mutational analysis. *Chem. Rev.* **97**, 1233–1250 (1997).
- Kwong, P. D. *et al.* Structures of HIV-1 gp120 envelope glycoproteins from laboratory-adapted and primary isolates. *Structure* **8**, 1329–1339 (2000).
- Luque, I. & Freire, E. A system for the structure-based prediction of binding affinities and molecular design of peptide ligands. *Methods Enzymol.* **295**, 100–127 (1998).
- Fouts, T. R., Binley, J. M., Trkola, A., Robinson, J. E. & Moore, J. P. Neutralization of the human immunodeficiency virus type 1 primary isolate JR-FL by human monoclonal antibodies correlates with antibody binding to the oligomeric form of the envelope glycoprotein complex. *J. Virol.* **71**, 2779–2785 (1997).
- Sullivan, N. *et al.* CD4-induced conformational changes in the human immunodeficiency virus type 1 gp120 glycoprotein: consequences for virus entry and neutralization. *J. Virol.* **72**, 4694–4703 (1998).
- Salzwedel, K., Smith, E. D., Dey, B. & Berger, E. A. Sequential CD4-coreceptor interactions in human immunodeficiency virus type 1 env function: soluble CD4 activates env for co-receptor-dependent fusion and reveals blocking activities of antibodies against cryptic conserved epitopes on gp120. *J. Virol.* **74**, 326–333 (2000).

- Moore, J. P., McKeating, J. A., Huang, Y. X., Ashkenazi, A. & Ho, D. D. Virions of primary human immunodeficiency virus type 1 isolates resistant to soluble CD4 (sCD4) neutralization differ in sCD4 binding and glycoprotein gp120 retention from sCD4-sensitive isolates. *J. Virol.* **66**, 235–243 (1992).
- Arthos, J. *et al.* Biochemical and biological characterization of a dodecameric CD4-Ig fusion protein. Implications for therapeutic and vaccine strategies. *J. Biol. Chem.* **277**, 11456–11464 (2002).
- Zhu, P., Olson, W. C. & Roux, K. H. Structural flexibility and functional valence of CD4-IgG2 (PRO 542): Potential for cross-linking human immunodeficiency type 1 envelope spikes. *J. Virol.* **75**, 6682–6686 (2001).
- Trautner, A., Schneider, J., Kiefer, H. & Karjalainen, K. Highly efficient neutralization of HIV with recombinant CD4-immunoglobulin molecule. *Nature* **339**, 68–70 (1989).
- Wyatt, R. *et al.* The antigenic structure of the human immunodeficiency virus gp120 envelope glycoprotein. *Nature* **393**, 705–711 (1998).
- Doyle, M. L. *et al.* Measurement of protein interaction bioenergetics: application to structural variants of anti-sCD4 antibody. *Methods Enzymol.* **323**, 207–230 (2000).
- Zhou, J. Y. & Montefiori, D. C. Antibody-mediated neutralization of primary isolates of human immunodeficiency virus type 1 in peripheral blood mononuclear cells is not affected by the initial activation state of the cells. *J. Virol.* **71**, 2512–2517 (1997).
- Kwong, P. D., Wyatt, R., Sattentau, Q. J., Sodroski, J. & Hendrickson, W. A. Oligomeric modeling and electrostatic analysis of the gp120 envelope glycoprotein of the human immunodeficiency virus. *J. Virol.* **74**, 1961–1972 (2000).
- Wyatt, R. *et al.* Analysis of the interaction of the human immunodeficiency virus type 1 gp120 envelope glycoprotein with the gp41 transmembrane glycoprotein. *J. Virol.* **71**, 9722–9731 (1997).
- Fung, M. S. *et al.* Identification and characterization of a neutralization site within the second variable region of human immunodeficiency virus type 1 gp120. *J. Virol.* **66**, 848–856 (1992).
- Rusche, J. R. *et al.* Antibodies that inhibit fusion of human immunodeficiency virus-infected cells bind a 24-amino acid sequence of the viral envelope, gp120. *Proc. Natl Acad. Sci. USA* **85**, 3198–3202 (1988).

Supplementary Information accompanies the paper on Nature's website (<http://www.nature.com/nature>).

Acknowledgements We thank M. Dybul for providing primary isolates pi102, pi104, pi202 and pi204; M. A. Gawinowicz for matrix-assisted laser desorption/ionization-time of flight analysis; Z. Moodie and A. Palmer for help with error analysis; A. Fauci, S. Harrison, A. Miranker, R. Seder and L. Shapiro for discussions; and D. Dimitrov, B. Kwong, N. Letvin, J. Mascola, G. Nabel and Q. Sattentau for comments. This work was supported by grants from the National Institutes of Health and by a Center for AIDS Research grant to the Dana-Farber Cancer Institute. The Dana-Farber Cancer Institute is also the recipient of a Cancer Center Grant from the National Institutes of Health. Columbia University is a participant in a Center for AIDS Research. R.W. was a fellow of the American Foundation for AIDS Research and P.D.K. was a recipient of a Burroughs Wellcome Career Development award.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to P.D.K. (e-mail: pdkwong@nih.gov).

Asymmetric inheritance of centrosomally localized mRNAs during embryonic cleavages

J. David Lambert* & Lisa M. Nagy

Department of Molecular and Cellular Biology, University of Arizona, Tucson, Arizona 85721, USA

During development, different cell fates are generated by cell-cell interactions or by the asymmetric distribution of patterning molecules. Asymmetric inheritance is known to occur either through directed transport along actin microfilaments into one daughter cell^{1,2} or through capture of determinants by a region of the cortex inherited by one daughter^{3–5}. Here we report a third mechanism of asymmetric inheritance in a mollusc embryo. Different messenger RNAs associate with centrosomes in different cells and are subsequently distributed asymmetrically during division. The segregated mRNAs are diffusely distributed in the cytoplasm and then localize, in a microtubule-dependent

* Present address: Department of Genetics, Yale University School of Medicine, New Haven, Connecticut 06520, USA.

manner, to the pericentriolar matrix. During division, they dissociate from the core mitotic centrosome and move by means of actin filaments to the presumptive animal daughter cell cortex. In experimental cells with two interphase centrosomes, mRNAs accumulate on the correct centrosome, indicating that differences between centrosomes control mRNA targeting. Blocking the accumulation of mRNAs on the centrosome shows that this event is required for subsequent cortical localization. These events produce a complex pattern of mRNA localization, in which different messages distinguish groups of cells with the same birth order rank and similar developmental potentials.

During embryonic cleavage cycles in molluscs and related proto-stome phyla, most of the early cell divisions are asymmetric, producing cells that differ in size, subsequent cleavage pattern and eventual fate (Fig. 1). Embryological experiments implicate both cell lineage and cell signalling in the establishment of embryonic cell fates in molluscs^{6–9}. While examining the roles of conserved developmental patterning genes in the mollusc *Ilyanassa obsoleta*, we found that mRNAs for several patterning genes show marked changes in subcellular localization during cleavage stages. Here we focus on three such genes: *IoEve*, which encodes a protein involved in anterior–posterior axis specification in other organisms¹⁰; *IoDpp*, the orthologue of *dpp* (or bone morphogenetic protein 2 (BMP2) and *BMP4* in vertebrates), which encodes a secreted signalling molecule^{11,12}; and *IoTld*, which encodes a secreted protease that is predicted to be a modifier of Dpp signalling¹³. *IoDpp* and *IoTld* have a role in specifying head precursor cells after the 24-cell stage (our unpublished data) and *IoEve* seems to be involved in patterning a subset of cells along the animal–vegetal axis of the embryo. We have focused on the events that position these mRNAs during early cleavage cycles.

The mRNAs were localized to the centrosome during the early cleavage cycles, and this localization led to asymmetric inheritance during cleavage (Fig. 2). In the four-cell stage, *IoEve* mRNA was found in all four cells on a spherical structure that lies between the nucleus and the animal pole of the embryo (Fig. 2g). Similarly, at the

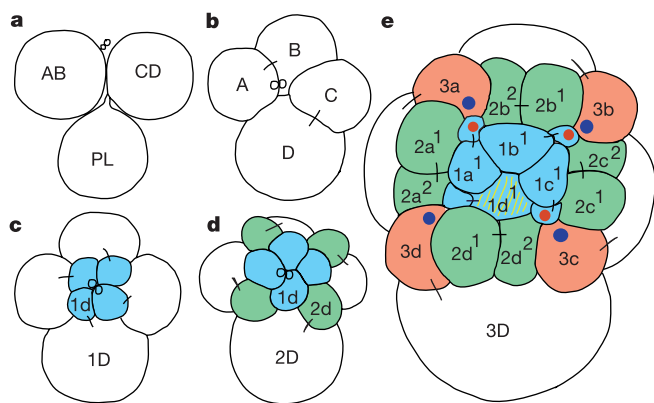


Figure 1 Diagram of *Ilyanassa* cleavage and patterns of centrosomally localized mRNA. **a, b**, The first two divisions are unequal and are accompanied by the production of a polar lobe (PL) at the vegetal pole, which results in a four-cell stage with one large and three smaller cells. These cells (macromeres) are the founder cells of the four lineages of the embryo, the A, B, C and D quadrants. **c–e**, In successive cleavage cycles, the macromeres function as stem cells, producing a stereotyped series of smaller cells towards the animal pole (micromeres). The four micromere cells produced by the macromeres in a given cleavage cycle are called a quartet. Quartet members have similar cleavage patterns and cell fates that differ from those of other quartets. This suggests that each quartet of micromeres receives similar patterning cues, although the mechanism is not known. The first quartet (1a–d) is blue (**c**), the second (2a–d) is green (**d**), and the third (3a–d) is orange (**e**). A summary of the patterns of centrosomally localized mRNAs described here (see Fig. 4) at the 24-cell stage is shown in **e**. *IoTld* is yellow, *IoEve* is red, and *IoDpp* is dark blue. Sister cells from the previous division are connected by a dash.

eight-cell stage, *IoDpp* mRNA was localized to a spherical structure located anticlockwise to the nucleus in all four of the macromere cells (Fig. 2a). These mRNA-binding structures were at the centre of the microtubule arrays in these cells and were rich in γ -tubulin and centrin (Fig. 2be). We therefore conclude that these mRNA-binding structures are the interphase centrosomes.

Comparisons of γ -tubulin, β -tubulin and centrin staining with the distribution of mRNA suggested that the mRNA is bound throughout the pericentriolar matrix (PCM). Transmission electron micrographs confirmed that a zone of PCM occupies a region of the size predicted from other markers (Fig. 2f), showing that the mRNA

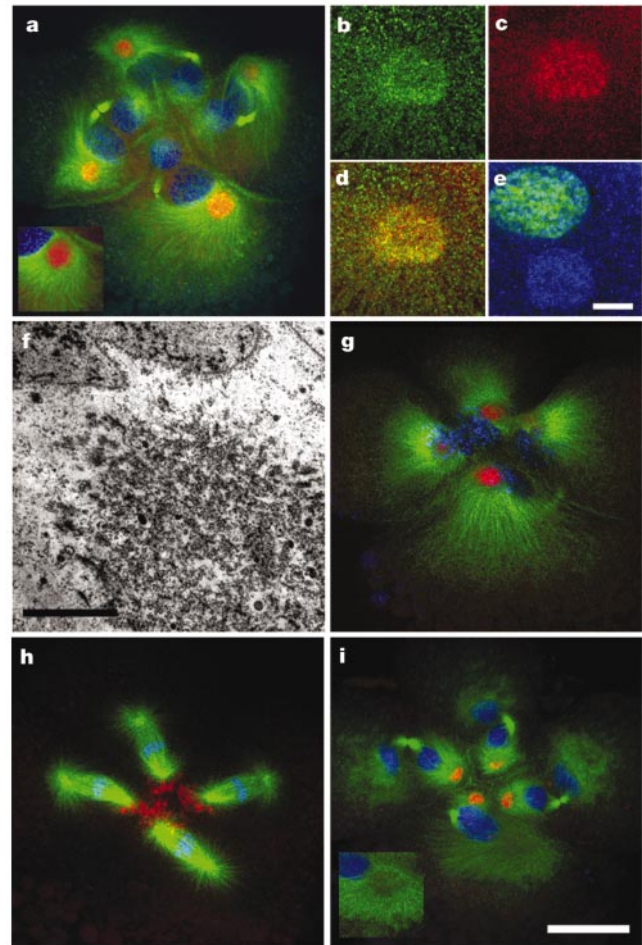


Figure 2 Localization of mRNAs to the centrosome and mRNA dynamics during cleavage. **a**, *In situ* hybridization with *IoDpp* mRNA to an eight-cell interphase embryo. The midbodies of the previous division are visible clockwise from each macromere nucleus (projection of confocal z series); inset, single section through the 1D centrosome (see Fig. 1b). **b–d**, γ -Tubulin (**b**) is bound to the same spherical structure as the *IoDpp* mRNA (**c**); merged in **d**. **e**, Centrin is enriched on a spherical structure located anticlockwise and vegetal to the macromere nuclei. **f**, Transmission electron micrograph of 1D nucleus (top left) and centrosome shows typical PCM composition of the size predicted by other markers. **g**, *In situ* hybridization with *IoEve* mRNA during early prophase of the four-cell stage (single confocal section, B and D macromere centrosomes in the plane of focus). **h, i**, *In situ* hybridization with *IoEve* mRNA during pro-metaphase at the transition to the eight-cell stage (**h**; projection) and in an early eight-cell embryo (**i**; projection). Inset, single section through the centrosome in 1D shows that *IoEve* is not present on this centrosome. mRNAs were visualized by HNPP/Fast Red precipitate (red; **a, c, d, g–i**), nuclei by DAPI (blue; **a, g–i**) or YOYO-1 (green; **e**), β -tubulin and γ -tubulin antibodies by Alexafluor-488-conjugated secondary antibodies (green; **a, b, d, g–i**), and monoclonal 20H5 antibodies^{26,27} against centrin with Cy5-conjugated secondary antibodies (blue; **e**). Animal views of the 1D centrosome are shown in **b–f**; animal views of whole embryos are shown in **a, g–i**. Scale bars, 5 μ m (**f**); 10 μ m (**e**); 50 μ m (**i**).

is localized throughout the PCM of the centrosome. By contrast, the distribution of the mRNA of a gene that is not predicted to be involved in patterning, the 40S ribosomal subunit, was distributed diffusely throughout the cytoplasm of all cells and not localized to centrosomes (data not shown).

The centrosomally localized mRNAs were partitioned into one daughter cell during division. For example, during prophase of the third cleavage, the *IoEve* mRNA moved to the granules in the region of the cortex of each macromere that was closest to the animal pole (Fig. 2h). At the same time, two small prophase asters were first detected in place of the large interphase centrosome, and these migrated into position for mitosis. During division, the cortically localized mRNA was segregated solely into the animal daughter cell, such that after cell division, the mRNA was present only on the centrosome in this cell (Fig. 2i). This movement between the interphase centrosome and the presumptive micromere cortex in mitosis was observed with all patterning mRNAs studied (see also Fig. 3b, c).

All of the centrosome-localized mRNAs that we examined showed these two discrete types of intracellular movement: an initial recruitment to particular interphase centrosomes, followed by a movement to a cortical region that would be inherited exclusively by one daughter cell in a given division. We investigated the cytoskeletal requirements of the observed mRNA movement by treating embryos with agents that specifically disrupt the cytoskeleton (Fig. 3). At the four-cell stage, *IoDpp* mRNA was distributed diffusely in the cytoplasm (Fig. 3a). During cytokinesis at the transition to the eight-cell stage, it became localized specifically to the centrosomes in macromeres, but not in micromeres (which are their sister cells), and the mRNA remained on the centrosomes during interphase (Fig. 3b). During prophase of the fourth division, the mRNA moved to a portion of the cortex that would be inherited by the second quartet cells (Fig. 3c).

Depolymerization of microtubules with nocodazole when *IoDpp* mRNA was still spread throughout the cytoplasm left most of the mRNA distributed diffusely in the cytoplasm, with some mRNA in disorganized filamentous structures (Fig. 3d). This was not due to a secondary loss of mRNA from the centrosome, because other

experiments showed that nocodazole treatment after localization had occurred did not disrupt the interphase pattern. Disrupting microfilament polymerization with cytochalasin B over the same time period did not prevent accumulation of mRNA on the centrosome (data not shown and Fig. 4f). Thus, an intact microtubule cytoskeleton is required for accumulation of mRNA on the centrosome.

We next investigated the basis of the movement of centrosome-bound mRNAs to the cortex during prophase. Inhibiting microtubule polymerization with nocodazole did not prevent movement of the mRNA (data not shown). But when actin filaments were induced to depolymerize with cytochalasin B just before and during the movement to the cortex, the mRNA remained near the nucleus and did not migrate to the cortex (Fig. 3e). Cytochalasin B treatment after the mRNA had moved to the cortex did not release

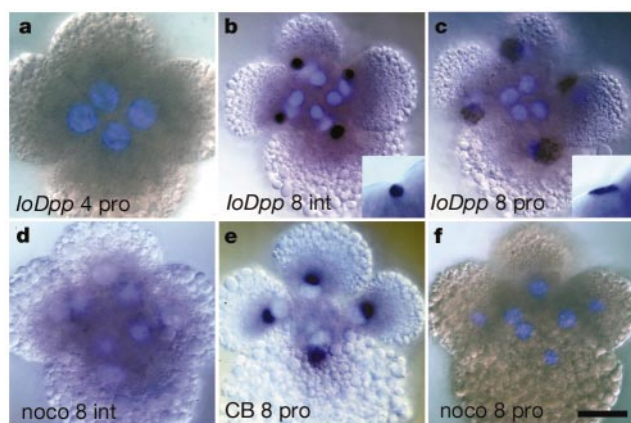


Figure 3 Cytoskeletal basis of mRNA localization. **a–c**, *In situ* hybridization with *IoDpp* mRNA at interphase of the four-cell stage (**a**), interphase of the eight-cell stage (**b**) and prophase of the eight-cell stage (**c**). Insets, lateral view of the 1D macromere centrosome (**b**); lateral view of *IoDpp* on the animal cortex of 1D (**c**). **d**, Embryo fixed 30 min after treatment with nocodazole (noco) during cytokinesis at the transition to eight-cell stage, when *IoDpp* mRNA was still spread around the cytoplasm (as in **a**). *IoDpp* mRNA was localized to the centrosome in controls (**b**). **e**, Treatment with cytochalasin B (CB) during interphase at the eight-cell stage, before *IoDpp* mRNA had moved to the cortex (compare with end-point control; **c**). **f**, Embryo treated with nocodazole as described in **d**, but fixed later during pro-metaphase. mRNAs are visualized by alkaline phosphatase precipitate (blue/black), nuclei by DAPI (light blue). Scale bar, 50 μ m.

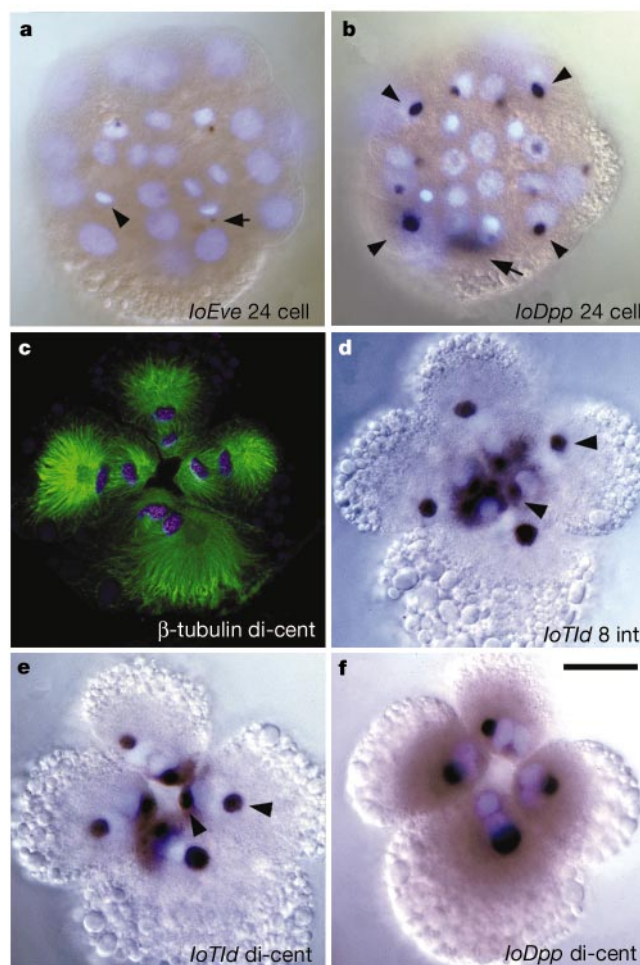


Figure 4 mRNAs are localized to specific subsets of cells during cleavage, and mRNAs are specifically targeted to particular centrosomes. **a**, *In situ* hybridization with *IoEve* mRNA at the 24-cell stage (arrow indicates $1c^2$, arrowhead indicates $1d^2$). **b**, *In situ* hybridization with *IoDpp* mRNA at the 24-cell stage. *IoDpp* is abundant on the centrosomes in 3a–d (arrowheads) and in a cloud above the nucleus of the 3D macromere (arrow). These patterns of localization are illustrated in Fig. 1e. **c**, **f**, **g**, Treatment with cytochalasin B during cytokinesis at the third cleavage cycle generates di-centrosomal embryos. **c**, Two microtubule-organizing centres (β -tubulin staining, green) and two nuclei (DAPI staining, blue) are seen in each of the four cells in cytochalasin-B-treated embryos (single confocal section). **d**, *In situ* hybridization with *IoTld* in a normal eight-cell embryo (arrowheads indicate centrosomes in the cells $1c$ and $1c$). **e**, *In situ* hybridization with *IoTld* in di-centrosomal embryos. Arrowheads indicate the pair of centrosomes in the single C quadrant cell. **f**, *In situ* hybridization with *IoDpp* in di-centrosomal embryos. mRNAs are visualized by alkaline phosphatase precipitate (blue/black), nuclei by DAPI (light blue). Scale bar, 50 μ m.

the mRNA (data not shown). Thus, actin filaments, but not intact microtubules, are required for attraction or docking at the cell periphery. We therefore propose that mRNAs that require segregation for their subsequent roles in embryonic patterning are loaded onto the centrosome by minus-end-directed transport along microtubules. This movement prepares them for their subsequent delivery by actin filaments to a region of the cell cortex that will be inherited by a particular daughter cell.

We next considered whether accumulation on the cortex would occur without previous localization to the centrosome. We blocked *IoDpp* mRNA accumulation on the centrosome (as shown in Fig. 3d) and examined *IoDpp* mRNA at the onset of mitosis, a stage when mRNA is localized on the cortex in control embryos. We did not observe any cortical *IoDpp* mRNA staining (Fig. 3f). This was not a consequence of arrest or delay after treatment, because the chromosomal arrangement and condensation indicated that macromere nuclei were in prophase and pro-metaphase as in control embryos, and also *IoDpp* mRNA was not cortically localized at later time points. Because nocodazole treatment after localization to the centrosome does not affect centrosomal localization or movement to the cortex, this result indicates that mRNA movement to the cortex requires its previous accumulation on the centrosome.

We followed *IoEve*, *IoDpp* and *IoTld* in later cleavage stages. All three were associated with centrosomes, and all three were sorted during cell division (Fig. 4a, b). By the 24-cell stage, *IoEve* became localized to three particular cells that give rise to the ciliated band of the larvae (the primary trochoblasts 1a–c²; Fig. 4a). *IoDpp* was segregated into the second and third quartet micromeres, but decayed in the second quartet cells after the production of the third quartet, resulting in a specific localization in the third quartet cells, as well as in one macromere (the '3D' cell; Fig. 4b). *IoTld* was loaded onto centrosomes in all of the first quartet cells, but persisted only in the dorsal cell 1d¹ (data not shown). These patterns of mRNA distribution show that several patterning genes are localized specifically to certain subsets of cells by centrosome-mediated sorting. In addition, the subsets of micromere cells that contain a given mRNA have the same birth order rank and have similar cell fates and cleavage patterns^{14–16}. Our results suggest that these groups may be distinguished by centrosome-mediated segregation of patterning molecules.

The patterns of mRNA accumulation on the centrosomes of different cells suggest that intrinsic differences between these centrosomes control message accumulation. To test this, we produced diploid cells that contained two interphase centrosomes by blocking cytokinesis at the transition to the eight-cell stage with cytochalasin B (Fig. 4c–f). *IoTld* mRNA, which was present on centrosomes of both the animal and vegetal daughter cells in controls, was present on both centrosomes in di-centrosomal cells (Fig. 4d, e), showing that in these cells the two centrosomes lie at predictable locations that correspond to the locations of the centrosomes in the two normal daughter cells. *IoDpp* mRNA, which was on only the vegetal daughter centrosomes in controls (Fig. 3b), was found on only vegetal centrosomes in di-centrosomal cells (Fig. 4f). Thus, mRNAs are specifically targeted to certain centrosomes, so that even within the same cytoplasm a message accumulates on the correct centrosome. This suggests that intrinsic differences between centrosomes result in accumulation of particular mRNAs on centrosomes in certain cells. Because localization to the centrosome is required for subsequent cortical localization and segregation, differences in mRNA recruitment between centrosomes could thus control the segregation of mRNAs in the next cleavage cycle.

The mechanisms of asymmetric inheritance of patterning molecules have been investigated in several other systems^{2–4}. In *I. obsoleta*, as in other systems such as the *Drosophila melanogaster* neuroblast divisions, segregated molecules bind to a particular region of the cortex before division^{3,5}. But the role that we describe

for the centrosome during asymmetric cell division—that is, the site where segregated mRNAs are localized before moving to the cortex—has not been reported. Although the prior localization of these molecules to the centrosome has not been described, it is apparently not unusual in this system: we observed it with several different mRNAs and in successive cleavage cycles. It may represent a general mechanism of embryonic patterning in this embryo. The role that we show for the centrosome is reminiscent of mechanisms of asymmetric cell division from other systems. In ascidians, regions of the cortex that bind localized mRNAs interact with centrosomes during asymmetric cell division^{17,18}. In the *Caenorhabditis elegans* embryo, degradation of PIE-1 protein on the anterior centrosome of P1 complements other mechanisms to segregate the protein to the posterior daughter^{19,20}. Our observations provide a link between the segregation of determinants and the cleavage apparatus in asymmetric cell divisions. Centrosomes replicate at each cell cycle in somatic cells, and the daughter centrosomes are generally considered to be equivalent. But evidence suggests that there can be functional differences between the two centrosomes in a dividing cell^{21–23}. Our work indicates that cells can exploit differences between centrosomes in different cells to target patterning molecules for asymmetric partitioning during division.

E. G. Conklin²⁴, working with the mollusc *Crepidula* in 1902, described a distinct portion of cytoplasm that surrounded the core mitotic centrosome during interphase in the macromeres. This structure moved to the cortex before division and was inherited in its entirety by the micromeres. He called this structure the 'sphere', to distinguish it from the core mitotic centrosome. Our results corroborate Conklin's observations and show that the sphere is a part of the interphase centrosome, and that it contains mRNAs for several developmental patterning genes. These results suggest that, during early cleavages in mollusc embryos, the centrosome is the site of assembly for a package of molecules that is then delivered to one daughter cell, and that this mechanism distinguishes sets of equivalent cells during cleavage. This mechanism may operate in other systems where a stem cell produces a stereotyped series of daughter cells with different fates. □

Methods

RNA *in situ* hybridization and immunohistochemistry

Fragments of complementary DNAs were cloned by polymerase chain reaction (PCR) with degenerate primers, followed by rapid amplification of cDNA ends with PCR to obtain larger fragments for *in situ* hybridization. For colocalization of mRNAs and the microtubule skeleton cytoskeleton, and for most standard *in situ* hybridizations, we fixed embryos in PEM (100 mM PIPES, pH 6.9, 10 mM EGTA and 1 mM MgSO₄) with 8% paraformaldehyde, 100 mM sucrose and 0.1% Triton X-100. For the *in situ* hybridizations shown in Fig. 4a, b, embryos were fixed in 3.7% formalin in 90% filtered artificial sea water. For all *in situ* hybridizations, embryos were pretreated for 10 min in 2% acetic anhydride in TEA buffer and prehybridized for 3 h at 63 °C in Hyb solution (5 × SSC, 1 × Denhart's, 50% formamide, 1% Tween 20, 100 µg ml⁻¹ heparin and 100 µg ml⁻¹ yeast rRNA plus tRNA), hybridized with 1–10 ng ml⁻¹ digoxigenin-labelled probe for 12–18 h, washed three times over 2 h with Hyb solution at 63 °C. Non-fluorescent *in-situ*s were detected by nitro blue tetrazolium/5-bromo-4-chloro-3-indolyl phosphate precipitation. For colocalization with the microtubule cytoskeleton, samples were incubated at 23 °C overnight with monoclonal mouse antibodies against β-tubulin (Developmental Studies Hybridoma Bank) plus antibodies against digoxigenin conjugated to alkaline phosphatase. Alkaline phosphatase detection of hybridized probes was carried out with HNPP/Fast Red substrate (Boehringer-Mannheim) as described²⁵, followed by overnight incubation at 23 °C with secondary antibodies against mouse IgG. We carried out γ-tubulin staining with a rabbit polyclonal antibody generated against an epitope found in human γ-tubulin (Sigma). Centrin staining was done with two mouse monoclonals (20H5 and 11B2) generated against the *Chlamydomonas reinhardtii* centrin protein, both of which recognize centrin in animal cells^{26,27}. Animal maintenance, embryo collection, antibody staining, YOYO-1 (Molecular probes) and 4',6-diamidino-2-phenylindole dihydrochloride (DAPI) staining and all other steps were done as described²⁸.

Transmission electron microscopy

Embryos were fixed for 3–6 h in 2% glutaraldehyde and 95% artificial sea water, and then postfixed in 2% OsO₄ for 1–2 h.

Inhibitor experiments

We made solutions of the inhibitors in dimethyl sulphoxide at 1,000 times the final concentrations. Nocodazole was used at 1 µg ml⁻¹ (Fig. 3d) or 10 µg ml⁻¹ (in all other

nocodazole assays); both concentrations were sufficient to block cleavage and depolymerize essentially all microtubules, as judged by staining with antibodies against β -tubulin. We used cytochalasin B at $20 \mu\text{g ml}^{-1}$, a concentration sufficient to block cleavage and to depolymerize essentially all actin filaments, as judged by phalloidin staining.

Received 12 February; accepted 10 October 2002; doi:10.1038/nature01241.

1. Takizawa, P. A., Sil, A., Swedlow, J. R., Herskowitz, I. & Vale, R. D. Actin-dependent localization of an RNA encoding a cell-fate determinant in yeast. *Nature* **389**, 90–93 (1997).
2. Long, R. M. *et al.* Mating type switching in yeast controlled by asymmetric localization of ASH1 mRNA. *Science* **277**, 383–387 (1997).
3. Spana, E. P. & Doe, C. Q. The Prospero transcription factor is asymmetrically localized to the cell cortex during neuroblast mitosis in *Drosophila*. *Development* **121**, 3187–3195 (1995).
4. Boyd, L., Guo, S., Levitan, D., Stinchcomb, D. T. & Kemphues, K. J. PAR-2 is asymmetrically distributed and promotes association of P granules and PAR-1 with the cortex in *C. elegans* embryos. *Development* **122**, 3075–3084 (1996).
5. Lu, B. W., Ackerman, L., Jan, L. Y. & Jan, Y. N. Modes of protein movement that lead to the asymmetric localization of partner of numb during *Drosophila* neuroblast division. *Mol. Cell* **4**, 883–891 (1999).
6. Crampton, H. E. Experimental studies on gastropod development. *Roux's Arch. EntwMech. Org.* **3**, 1–19 (1896).
7. Clement, A. C. Experimental studies on germinal localization in *Ilyanassa*. I. The role of the polar lobe in determination of the cleavage pattern and its influence in later development. *J. Exp. Zool.* **132**, 427–446 (1952).
8. Wilson, E. B. Experimental studies in germinal localization. II Experiments on the cleavage-mosaic in patella and dentalium. *J. Exp. Zool.* **1**, 197–268 (1904).
9. Clement, A. C. Development of *Ilyanassa* following the removal of the D macromere at successive cleavage stages. *J. Exp. Zool.* **149**, 193–216 (1962).
10. Macdonald, P. M., Ingham, P. & Struhl, G. Isolation, structure, and expression of *even-skipped*—a 2nd pair-rule gene of *Drosophila* containing a homeo box. *Cell* **47**, 721–734 (1986).
11. Wharton, K. A., Ray, R. P. & Gelbart, W. M. An activity gradient of Decapentaplegic is necessary for the specification of dorsal pattern elements in the *Drosophila* embryo. *Development* **117**, 807–822 (1993).
12. Holley, S. A. *et al.* A conserved system for dorsal–ventral patterning in insects and vertebrates involving Sog and Chordin. *Nature* **376**, 249–253 (1995).
13. Shimell, M. J., Ferguson, E. L., Childs, S. R. & O'Connor, M. B. The *Drosophila* dorsal–ventral patterning gene *Tollid* is related to human bone morphogenetic protein-1. *Cell* **67**, 469–481 (1991).
14. Clement, A. C. Cell determination and organogenesis in molluscan development—reappraisal based on deletion experiments in *Ilyanassa*. *Am. Zool.* **16**, 447–453 (1976).
15. Render, J. Cell fate maps in the *Ilyanassa obsoleta* embryo beyond the third division. *Dev. Biol.* **189**, 301–310 (1997).
16. Sweet, H. C. Specification of first quartet micromeres in *Ilyanassa* involves inherited factors and position with respect to the inducing D macromere. *Development* **125**, 4033–4044 (1998).
17. Yoshida, S., Marikawa, Y. & Satoh, N. posterior end mark, a novel maternal gene encoding a localized factor in the ascidian embryo. *Development* **122**, 2005–2012 (1996).
18. Nishikata, T., Hibino, T. & Nishida, H. The centrosome-attracting body, microtubule system, and posterior egg cytoplasm are involved in positioning of cleavage planes in the ascidian embryo. *Dev. Biol.* **209**, 72–85 (1999).
19. Mello, C. C. *et al.* The PIE-1 protein and germline specification in *C. elegans* embryos. *Nature* **382**, 710–712 (1996).
20. Reese, K. J., Dunn, M. A., Waddle, J. A. & Seydoux, G. Asymmetric segregation of PIE-1 in *C. elegans* is mediated by two complementary mechanisms that act through separate PIE-1 protein domains. *Mol. Cell* **6**, 445–455 (2000).
21. Wu, X. Y. & Palazzo, R. E. Differential regulation of maternal vs. paternal centrosomes. *Proc. Natl Acad. Sci. USA* **96**, 1397–1402 (1999).
22. Bonaccorsi, S., Giansanti, M. G. & Gatti, M. Spindle assembly in *Drosophila* neuroblasts and ganglion mother cells. *Nature Cell Biol.* **2**, 54–56 (2000).
23. Piel, M., Meyer, P., Khodjakov, A., Rieder, C. L. & Bornens, M. The respective contributions of the mother and daughter centrioles to centrosome activity and behavior in vertebrate cells. *J. Cell Biol.* **149**, 317–329 (2000).
24. Conklin, E. G. Karyokinesis and cytokinesis in the maturation, fertilization and cleavage of Crepidula and other gasteropoda. *J. Acad. Nat. Sci. Philadelphia Ser. 2* **12**, 1–21 (1902).
25. Goto, S. & Hayashi, S. Cell migration within the embryonic limb primordium of *Drosophila* as revealed by a novel fluorescence method to visualize mRNA and protein. *Dev. Genes Evol.* **207**, 194–198 (1997).
26. Sanders, M. A. & Salisbury, J. L. Centrin plays an essential role in microtubule severing during flagellar excision in *Chlamydomonas reinhardtii*. *J. Cell Biol.* **124**, 795–805 (1994).
27. Paoletti, A., Moudjou, M., Paintrand, M., Salisbury, J. L. & Bornens, M. Most of centrin in animal cells is not centrosome-associated and centrosomal centrin is confined to the distal lumen of centrioles. *J. Cell Sci.* **109**, 3089–3102 (1996).
28. Lambert, J. D. & Nagy, L. M. MAPK signaling by the D quadrant embryonic organizer of the mollusc *Ilyanassa obsoleta*. *Development* **128**, 45–56 (2001).

Acknowledgements We thank J. Cooley, D. Bentley and J. Wandelt for technical help; J. Salisbury and G. Hermann for antibodies; D. Brower, C. Gregorio and G. Von Dassow for critically reading the manuscript; and R. Palazzo, M. Goulding, G. Lambert, E. Wilk and M. Gibson for technical advice and discussions. This work was supported by an National Science Foundation (NSF) Graduate Research Fellowship and an NSF Doctoral Dissertation Improvement Grant to J.D.L., and a grant from the NSF to L.M.N.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to L.M.N. (e-mail: lnagy@u.arizona.edu). Sequences have been deposited in GenBank under accession codes AF499914 (*IoDpp*), AF499912 (*IoEve*), AF499913 (*IoTld*) and AY151152 (*Io40S*).

Endocytosis-mediated downregulation of LIN-12/Notch upon Ras activation in *Caenorhabditis elegans*

Daniel D. Shaye* & Iva Greenwald†‡

* Departments of Genetics and Development and † Biochemistry and Molecular Biophysics, and ‡ Howard Hughes Medical Institute, Columbia University, College of Physicians and Surgeons, New York, New York 10032, USA

The coordination of signals from different pathways is important for cell fate specification during animal development. Here, we define a novel mode of crosstalk between the epidermal growth factor receptor/Ras/mitogen-activated protein kinase cascade and the LIN-12/Notch pathway during *Caenorhabditis elegans* vulval development. Six vulval precursor cells (VPCs) are initially equivalent but adopt different fates as a result of an inductive signal mediated by the Ras pathway and a lateral signal mediated by the LIN-12/Notch pathway¹. One consequence of activating Ras is a reduction of LIN-12 protein in P6.p (ref. 2), the VPC believed to be the source of the lateral signal. Here we identify a 'downregulation targeting signal' (DTS) in the LIN-12 intracellular domain, which encompasses a di-leucine-containing endocytic sorting motif. The DTS seems to be required for internalization of LIN-12, and on Ras activation it might mediate altered endocytic routing of LIN-12, leading to downregulation. We also show that if LIN-12 is stabilized in P6.p, lateral signalling is compromised, indicating that LIN-12 downregulation is important in the appropriate specification of cell fates *in vivo*.

The six VPCs, consecutively numbered P3.p–P8.p, have the potential to adopt one of three fates, 1°, 2° or 3°, which can be distinguished by cell lineage and marker gene expression (see Fig. 1). In wild-type hermaphrodites, P3.p–P8.p always adopt the same pattern of fates (3°, 3°, 2°, 1°, 2° and 3°, respectively) as a result of inductive and lateral signalling events (reviewed in ref. 1). The inductive signal, LIN-3, produced by the anchor cell of the gonad activates the receptor tyrosine kinase LET-23 and a canonical Ras/mitogen-activated protein kinase (MAPK) cascade in P6.p to promote the 1° fate. An unknown lateral signal, believed to be produced by P6.p, activates LIN-12 in P5.p and P7.p to promote the 2° fate. Loss of SUR-2, a Ras target that is a component of the Mediator transcription factor complex³, abrogates lateral signalling⁴; thus Ras activation might control the transcription of the lateral signal and/or of factors required to activate this signal.

Ras activation also causes downregulation of a LIN-12::green-fluorescent-protein (GFP) reporter in P6.p and its descendants². However, a *lin-12::lacZ* transcriptional reporter is expressed continuously in the VPCs and their daughters⁵, suggesting that transcriptional control of *lin-12* is not the mechanism by which downregulation of LIN-12 is accomplished, in contrast to another paradigmatic LIN-12-mediated cell fate decision in *C. elegans*⁶. Here we explore two related issues, the mechanism of post-transcriptional downregulation of LIN-12 in response to Ras activation, and the role of this downregulation in VPC fate specification.

We hypothesized that there would be a DTS within LIN-12. To identify this signal, we used the *egl-17* promoter (*egl-17p*), which is activated in P6.p in response to the inductive signal⁷, to express GFP-tagged LIN-12 fragments in P6.p and its descendants. Extra-chromosomal arrays carrying such constructs were marked with *egl-17p::LacZ*, and transgenic hermaphrodites were stained with anti-GFP and anti-LacZ antibodies at the time of vulval induction (see Methods). LacZ will be expressed in P6.p and its daughters after

inductive signalling; such LacZ-positive hermaphrodites were scored for the presence of GFP. If the LIN-12 fragment being assayed carried a DTS, GFP staining should not be detected; if the fragment lacked a DTS, then GFP should be detected (Fig. 2a).

We first assayed *egl-17p::LIN-12(+):GFP* and found that it is efficiently downregulated (Figs 2c and 3a). In contrast, GFP targeted to the plasma membrane by the SEL-1 signal sequence⁸ and a synthetic transmembrane domain (sTM)⁹ is not downregulated (Figs 2b and 3b). These results demonstrate the efficacy of our approach and show that the LIN-12 protein contains a sequence mediating downregulation. We found that the LIN-12 extracellular domain is not necessary for downregulation, but GFP-tagged intracellular domain (LIN-12(intra)::GFP), which localizes mostly to the nucleus (data not shown), was not efficiently downregulated (Fig. 3a). Targeting LIN-12(intra)::GFP to the membrane by using the sTM permits efficient downregulation (Fig. 3a). We made subsequent deletions in sTM::LIN-12(intra)::GFP and identified a 90-amino-acid region that is necessary and sufficient to promote downregulation (LIN-12 amino acids 940I to 1029E, referred to as r1; Fig. 3a). Deletion of 940I to 956R or 972A to 1029E within r1 did not affect downregulation (data not shown). In contrast, deletion of the sequence 957KNHQSITSSQHSLE971 prevented efficient downregulation of sTM::LIN-12(intra)::GFP (Fig. 3a). We refer to this discrete 15-amino-acid sequence as the DTS (Fig. 3b).

The DTS contains two adjacent leucine residues and several serines that are conserved in all known nematode LIN-12/Notch proteins (see Supplementary Information 1). 'Di-leucine motifs' are well-characterized sorting signals that usually take the form (–)(2–4)XLL, where (–) is often an acidic residue or phosphoserine, although basic residues have also been seen at this position¹⁰, and X

is usually a polar or bulky residue¹¹. Functional motifs that contain M, V or I instead of L have also been found^{10,11}. Di-leucine motifs regulate different aspects of protein trafficking, including the constitutive or ligand-stimulated internalization of transmembrane receptors^{12–14}, and the routing of proteins within the endocytic and/or secretory pathways^{10,15}. Different residues within the same di-leucine motif might modulate different aspects of motif activity (namely internalization versus routing)¹⁰, and the activity of di-leucine motifs that contain serines might be regulated by phosphorylation¹². In several cases the di-leucine motif leads to routing of the protein to lysosomes for degradation^{13,15}.

Mutating the two leucine residues of the DTS to two alanines disrupted downregulation of sTM::LIN-12(intra)::GFP (Fig. 3a). Because downregulation of GFP-tagged LIN-12 fragments requires both membrane association and a di-leucine motif, we infer that the mechanism of downregulation involves increased internalization and/or altered endocytic routing of LIN-12 when Ras is activated. We elaborate on this point below.

To investigate the role of Ras-promoted downregulation of LIN-12, we deleted the DTS from an otherwise intact LIN-12::GFP protein in a *lin-12* genomic context (*lin-12g*) that contains all regulatory sequences necessary for phenotypic rescue of a *lin-12(0)* allele¹⁶. We first assessed whether *lin-12g::LIN-12(ΔDTS)::GFP* could function normally in non-vulval cells, where post-transcriptional downregulation does not occur. Extra-chromosomal arrays carrying *lin-12g::LIN-12(ΔDTS)::GFP* are able to rescue the hallmark non-vulval defects of *lin-12(0)* hermaphrodites (sterility and an extra anchor cell) as efficiently as arrays carrying *lin-12g::LIN-12(+):GFP*; these arrays also do not cause any hypermorphic or dominant-negative non-vulval phenotypes in a *lin-12(+)* background (see Methods and Supplementary Information 2). We conclude that LIN-12(ΔDTS)::GFP functions like LIN-12(+):GFP in situations where DTS-mediated post-transcriptional downregulation should not be relevant.

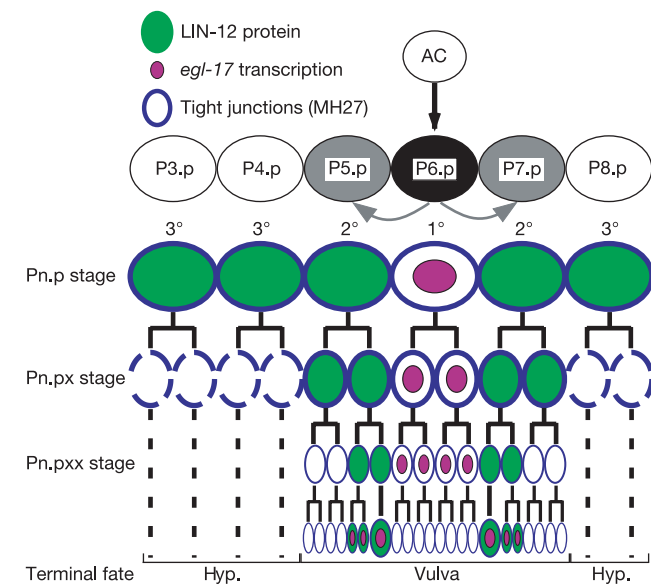


Figure 1 VPC fate specification and cell fate markers. In the third larval stage, an inductive signal (black arrow) from the gonadal anchor cell (AC) and lateral signalling (grey arrows) among the VPCs impose a 3°-3°-2°-1°-2°-3° pattern of cell fates. The three fates can be distinguished by cell lineage and the expression pattern of three markers, *egl-17::lacZ* (ref. 7) (purple), LIN-12::GFP² (green) and the cell-junction marker MH27 (ref. 28) (blue). Before inductive signalling, all cells express MH27 and LIN-12::GFP (not shown). The 1° fate is characterized by expression of MH27 in all descendants, expression of *egl-17* from the time of induction until the Pn.pxx stage, and post-transcriptional downregulation of LIN-12::GFP. The 2° fate is characterized by expression of MH27 in all descendants, expression of *egl-17* in the terminal stage, and continued expression of LIN-12::GFP in some 2° descendants. The 3° fate is characterized by a single round of division, followed by loss of all markers and fusion with the hypodermal syncytium (Hyp.; dashed lines). Note that loss of LIN-12::GFP from the 3° lineage and certain 2° descendants at the Pn.pxx stage is due to transcriptional regulation⁵.

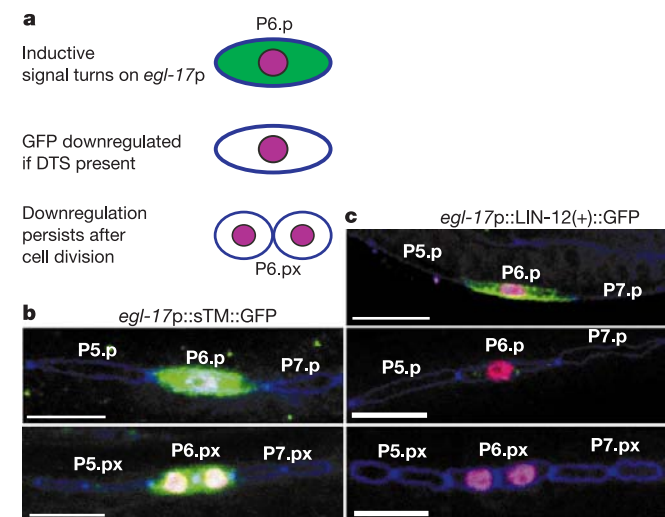


Figure 2 Strategy for defining the LIN-12 downregulation targeting signal (DTS). In photomicrographs here and in Fig. 4, anterior is to the left and the scale bar represents 10 μm. Views are ventral unless otherwise stated. Green fluorescent protein (GFP; green), LacZ (purple) and MH27 (blue) were detected by immunofluorescence (see Methods). **a**, General approach. *egl-17* is expressed in response to the inductive signal before LIN-12::GFP downregulation²⁹. Nuclear-localized LacZ (purple) and GFP-tagged LIN-12 fragments (green) are under the control of *egl-17p*. **b**, Membrane-targeted GFP (sTM::GFP) is detected continuously in P6.p and its daughters (P6.px). **c**, LIN-12(+):GFP is initially visible in P6.p, but is downregulated before the P6.p division and is not visible in the P6.px stage and thereafter. The top panel is a lateral view, highlighting the apical (towards the bottom) accumulation of LIN-12(+):GFP before downregulation. The other panels show ventral views of P6.p (middle) and P6.px (bottom) after downregulation.

We next examined the expression and function of LIN-12(Δ DTS)::GFP in the VPCs. If the DTS is required for downregulation, and if downregulation is functionally important for VPC fate specification, then persistent LIN-12(Δ DTS)::GFP in P6.p should cause a gain-of-function VPC fate defect. We therefore integrated extrachromosomal arrays carrying *lin-12g*::LIN-12(Δ DTS)::GFP in a *lin-12(+)* background (see Methods) and examined their properties. As expected, LIN-12(Δ DTS)::GFP was not efficiently downregulated in P6.p and its descendants in a *lin-12(+)* background (Fig. 4f, h). In a *lin-12(0)* background, LIN-12(Δ DTS)::GFP persists in P6.p in a lower proportion of hermaphrodites (data not shown). We therefore focus here on experiments performed in a *lin-12(+)* background, where we can best assess the consequences of persistent LIN-12 in P6.p.

We noted a clear difference between the subcellular localization of LIN-12(+>::GFP and LIN-12(Δ DTS)::GFP in the VPCs, where Ras is not active. LIN-12(+>::GFP accumulates at the apical membrane (Fig. 4c, d) and in puncta at or below the membrane (Fig. 4a–d) reminiscent of endocytic vesicles observed in *C. elegans* oocytes¹⁷, coelomocytes¹⁸ and synaptic vesicles in neurons¹⁹. In fact, simultaneous staining with an early endocytic marker showed that at least some of the LIN-12(+>::GFP puncta are early endosomes (see Supplementary Information 3). In contrast, LIN-12(Δ DTS)::GFP protein accumulates mostly at the apical membrane (Fig. 4h, i), with no significant concentration in puncta (Fig. 4f–i). These results suggest that the DTS is required for internalization of LIN-12, and that activation of Ras can only downregulate LIN-12 protein that is able to be internalized, possibly by regulating the rate of internalization and/or the routing of endocytosed LIN-12.

To test the role of downregulation, we assessed VPC fates in hermaphrodites carrying *lin-12g*::LIN-12(Δ DTS)::GFP (see Fig. 1 legend and Fig. 4e for VPC fate criteria). We found that P6.p adopts the 1° fate (Fig. 4e, j and Supplementary Information 4), suggesting that downregulation of LIN-12 is not necessary for adoption of the 1° fate. However, in hermaphrodites carrying *lin-12g*::LIN-12(Δ DTS)::GFP, P5.p and P7.p fail to adopt the 2° fate, and instead adopt the 3°, non-vulval, fate (Fig. 4j and Supplementary Information 4). This observation suggests that lateral signalling is impaired when LIN-12 is not downregulated (Fig. 4k). Simple elevation of LIN-12 levels due to expression from the transgene does not account for the failure of downregulation or the failure of lateral signalling: loss of SEL-10, which targets LIN-12/Notch proteins for turnover²⁰, leads to a high initial level of LIN-12(+>::GFP in all six VPCs but does not impair the downregulation of LIN-12(+>::GFP in P6.p (data not shown) or lateral signalling²⁰.

Under the control of *lin-12g*, LIN-12(Δ DTS)::GFP is expressed in all VPCs. When we used *egl-17p* instead of *lin-12g* to restrict the expression of LIN-12(Δ DTS)::GFP only to P6.p, we found that P5.p and P7.p failed to adopt the 2° fate and instead adopted the 3° fate. This result suggests that the defect in lateral signalling results from persistent LIN-12 in the signalling cell (Fig. 4k), and that downregulation of LIN-12 is required for the generation and/or activation of the lateral signal. Protein–protein interactions between LIN-12 and its ligand in the same cell might be inhibitory^{21,22}; alternatively, if persistent LIN-12 becomes activated by ligand, it might repress ligand gene expression^{6,23,24}. Regardless of the mechanism, the fact that persistent LIN-12 inhibits lateral signalling demonstrates a novel mode of cross-talk between the Ras and LIN-12/Notch

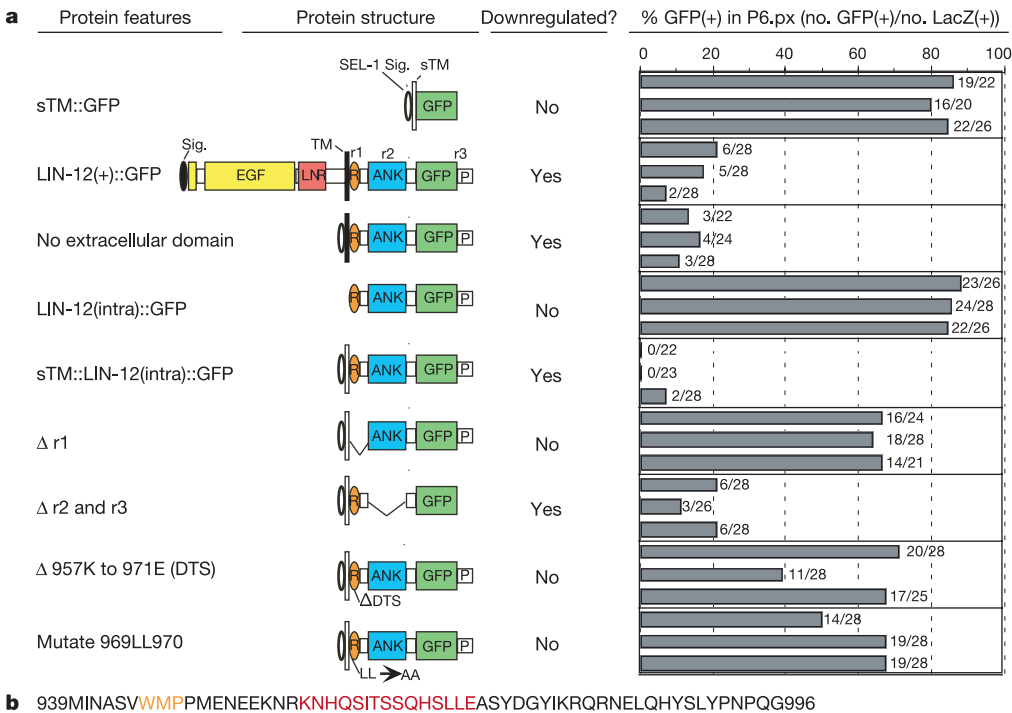


Figure 3 Identification of the LIN-12 downregulation targeting signal (DTS). **a**, Ability of different LIN-12 fragments to downregulate green fluorescent protein (GFP). To ensure that P6.p has been induced and that there has been sufficient time for downregulation, we score for downregulation at the Pn.px stage. Individual bars in the graph represent independent transgenic lines (see Supplementary Information 5 for full genotypes). LIN-12(+>::GFP represents the full-length LIN-12 protein, with hallmark regions diagrammed as follows: Sig., signal sequence; EGF, region containing repeated epidermal-growth-factor motifs; LNR, region containing three LIN-12/Notch repeat motifs; TM, transmembrane domain; R, region containing the RAM domain, which mediates interaction with the transcription factor LAG-1; ANK, region containing repeated

ankyrin motifs; P, PEST domain. In all constructs containing LIN-12 sequences, GFP was inserted between amino acids 1329R and 1340R, leading to a small deletion that did not impair LIN-12 function². The intracellular domain (intra) was divided into three regions for this analysis: r1 (from the end of the predicted transmembrane domain to the beginning of the first ANK motif), r2 (containing all ANK motifs), and r3 (from the end of the ANK region to the carboxy terminus). In some constructs, the LIN-12 signal sequence and TM were replaced with the SEL-1 signal sequence (white oval) and with a synthetic TM (white rectangle). **b**, LIN-12 RAM domain with the DTS in red and conserved amino acids required for transcription-factor binding in orange³⁰.

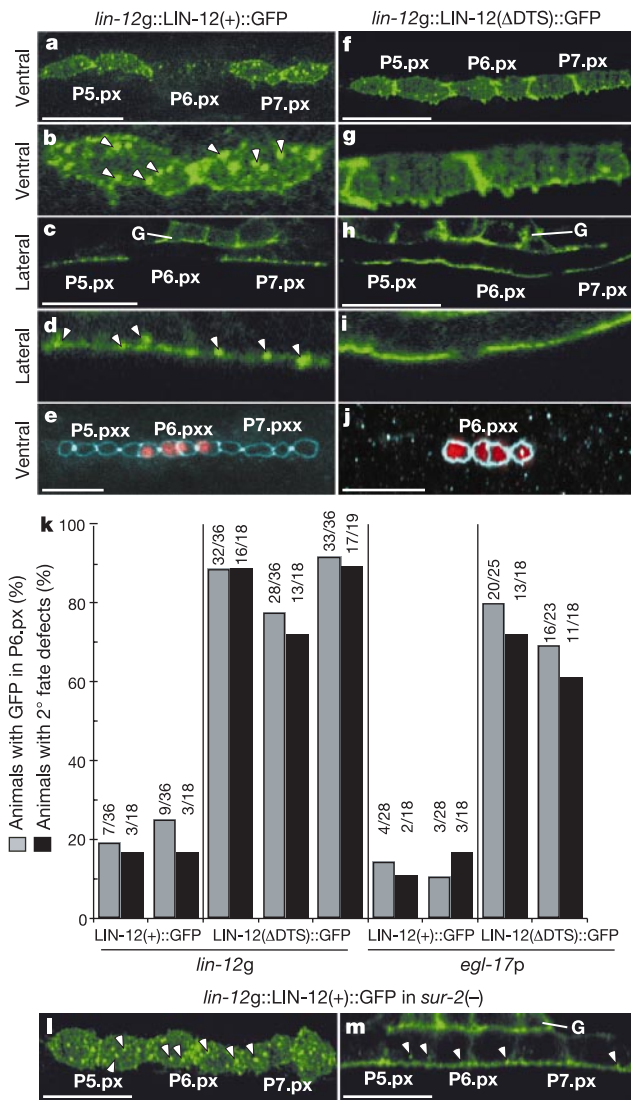


Figure 4 Failure to downregulate LIN-12, either by DTS deletion or mutations in *sur-2*, inhibits lateral signalling. Green fluorescent protein (GFP; green), LacZ (red) and MH27 (teal) were detected by immunofluorescence (see Methods). Ventral views show cell surfaces. Lateral views show cross-sections through cells, with the apical membrane towards the bottom and gonadal staining (G) seen above the cells. **a, c, f, h**, LIN-12(+):GFP is downregulated in P6.px, whereas LIN-12(ΔDTS):GFP is not. **b, d, g, i**, P7.px, 2.2 × zoom. In P5.px and P7.px, LIN-12(+):GFP is found in many puncta (arrowheads) reminiscent of endocytic vesicles, whereas LIN-12(ΔDTS):GFP is not. **e, j**, In hermaphrodites expressing LIN-12(+):GFP, all six VPCs are specified normally, whereas in hermaphrodites expressing LIN-12(ΔDTS):GFP, P5.p and P7.p do not adopt the 2° fate and instead adopt the 3° fate. Cell fate criteria are described in Fig. 1; in brief, 1° generate granddaughters that express *egl-17::LacZ* and MH27; 2° generate granddaughters that express MH27; 3° daughters do not divide or express MH27. **k**, Failure to downregulate LIN-12 in P6.p correlates with loss of lateral signalling. Each pair of bars represents results for a single transgenic line. The grey bar represents persistence of LIN-12::GFP in P6.px, and the black bar represents a lateral signalling defect (failure of P5.p and/or P7.p to adopt a 2° fate). For *lin-12g* constructs, integrated transgenes were used; for *egl-17p* constructs, extrachromosomal arrays were used (see Supplementary Information 5 for full genotypes). **l, m**, *SUR-2* is required for LIN-12 downregulation. Ventral (**l**) and lateral (**m**) views of *sur-2(ku9); arls41* hermaphrodites, showing failure to downregulate LIN-12(+):GFP in P6.p descendants. *sur-2(ku9)* is a strong loss-of-function allele of *sur-2* and exhibits a highly penetrant lateral signalling defect⁴. *arls41* is an integrated array carrying *lin-12g::LIN-12(+):GFP*. Of 36 *sur-2(ku9); arls41* hermaphrodites, 33 (92%) displayed a failure of LIN-12(+):GFP downregulation in P6.p descendants. We note that LIN-12(+):GFP accumulates in puncta (arrowheads) in *sur-2(ku9)* as it does in wild-type hermaphrodites (**a–d**).

pathway that is important for proper cell fate specification *in vivo*.

The gene *sur-2* encodes a component of the Mediator complex, which activates transcription in response to Ras/MAPK signalling in mammalian cells³. SUR-2 also seems to be activated by the Ras pathway in *C. elegans*, and hermaphrodites lacking *sur-2* activity display a failure in lateral signalling⁴. In *sur-2(-)* hermaphrodites, LIN-12(+):GFP is not downregulated (Fig. 4l, m). This result suggests that the failure of lateral signalling in *sur-2(-)* hermaphrodites might result at least in part from the failure to downregulate endogenous LIN-12(+). Furthermore, in *sur-2(-)* hermaphrodites, as in the wild type, LIN-12(+):GFP accumulates in puncta, suggesting that SUR-2/Mediator-promoted transcription is not necessary for the initial internalization, but instead might affect the rate of internalization and/or routing of endocytosed LIN-12(+):GFP.

Our model for cross-talk between the Ras and LIN-12 pathways in P6.p is summarized in Fig. 5. LIN-12, like other transmembrane proteins¹⁴, seems to be constitutively internalized and might be routed to recycling endosomes or to lysosomes (Fig. 4a). We propose that Ras activation leads to transcription of at least one gene whose product regulates the rate of internalization and/or

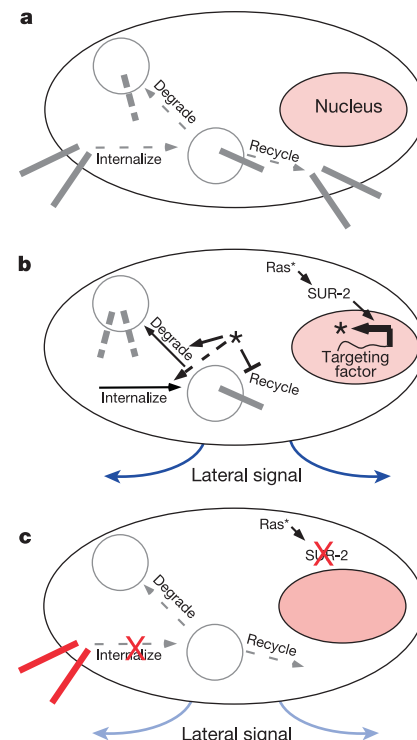


Figure 5 Model for the mechanism and role of endocytosis-mediated downregulation of LIN-12 in response to Ras activation in P6.p. **a**, Before induction. LIN-12 (bar) seems to display a degree of constitutive internalization (see Fig. 3b, d and Supplementary Information 3) and is possibly routed to recycling endosomes and/or to lysosomes. We represent this presumed constitutive endocytic trafficking in grey. **b**, After induction. The inductive signal is believed to lead to the production of a lateral signal (blue arrows) that activates LIN-12 in neighbouring VPCs. Our data suggest that activation of Ras by the inductive signal leads to SUR-2-mediated transcription of a factor (asterisk) that targets LIN-12 for downregulation. This unknown factor does not seem to be required for the initial internalization of LIN-12, but might enhance the rate of internalization, preferentially route LIN-12 for degradation, and/or inhibit recycling. Our data do not address possible additional effects of activating the Ras/MAPK pathway, such as direct phosphorylation of LIN-12. **c**, Persistence of LIN-12 (red bar) in P6.p can be achieved by deleting the DTS from LIN-12 (which inhibits internalization) or by the loss of *sur-2* activity. Failure to downregulate LIN-12 is correlated with an inhibition of lateral signalling. These observations suggest that one aspect of the Ras and SUR-2-mediated production or activation of the lateral signal might be the downregulation of LIN-12 in P6.p.

subcellular routing of LIN-12, leading to its degradation (Fig. 5b). Internalization (and perhaps, in addition, endocytic routing) of LIN-12 is mediated by the DTS: when the DTS is removed, LIN-12 accumulates in the apical membrane of P6.p. Persistence of LIN-12, achieved by deleting the DTS or by removing *sur-2* activity, is correlated with inhibition of lateral signalling (Fig. 5c), demonstrating that downregulation is functionally important.

The novel model of crosstalk between the Ras and LIN-12/Notch pathways that we have described here may be conserved. Vertebrate Notch proteins contain a highly conserved potential di-leucine-like motif (1822KKFRFEPPVVL1832 in human Notch1) that, like the LIN-12 DTS, lies between the transmembrane domain and the first ankyrin motif, and furthermore the large intracellular domain of notch proteins contains additional potential endocytic motifs. Whether these motifs function as such, and whether their function is regulated by Ras or other signals, can only be answered through experiments in other systems. We note that observations of Wingless endocytosis has led to the speculation that Ras might modulate the endocytic routing of Wingless bound to its receptor Frizzled2 via a di-leucine motif on Frizzled2, dependent on the transcription of an unknown factor in response to epidermal growth factor receptor/Ras/MAPK signalling²⁵. If this proves to be so, the future identification of the putative factors that recognize the targeting signals of LIN-12/Notch and Frizzled2 will reveal whether Ras works through a common or distinct mechanism to modulate the activity of these different receptors. □

Methods

Transgenic strains

Plasmids and strains are listed in Supplementary Information 5. Extrachromosomal arrays were generated by germline injections²⁶. *egl-17p*-driven constructs were injected at a concentration of 20 µg ml⁻¹ or *lin-12g*-driven constructs were injected at 5 µg ml⁻¹, along with plasmids *egl-17p::lacZ* (20 µg ml⁻¹ pNH291; ref. 7), *unc-4(+)* (30 µg ml⁻¹ pNC4.21; ref. 27) and pBluescript (Stratagene) to bring total DNA concentration to 100 µg ml⁻¹. Mixes containing *egl-17p*-driven constructs were injected into *unc-4(e120)* animals. Mixes containing *lin-12g*-driven constructs were injected into *unc-4(e120)*; *lin-12(n941)/qC1* animals. Non-Unc F₁ progeny were picked separately to establish independent extrachromosomal lines.

The ability of *lin-12g*-driven constructs to rescue *lin-12(n941)* sterility was assessed: 3/4 *lin-12g::LIN-12(ΔDTS)::GFP* extrachromosomal arrays (*arEx298*, *arEx299* and *arEx300*) and 5/6 *lin-12g::LIN-12(+):GFP* extrachromosomal arrays (we kept three, *arEx301*, *arEx302* and *arEx303*) permitted *lin-12(n941)* homozygotes to be fertile. The ability of these extrachromosomal arrays to rescue other *lin-12(n941)*-associated phenotypes are discussed in Supplementary Information 2.

We placed extrachromosomal arrays carrying *lin-12g*-driven constructs into a *lin-12(+)* background, and obtained integrants by using standard methods²⁶. We kept three independent integrated lines derived from *arEx299* (*arIs77*, *arIs79* and *arIs80*), one line from *arEx301* (*arIs82*) and one line from *arEx302* (*arIs84*). All integrated lines were backcrossed four times to *unc-4(e120)* before further analysis.

Immunofluorescence

Synchronization and fixation of animals are described in ref. 2. Synchronized animals were grown at 25 °C to ensure optimal expression from the *egl-17* promoter⁷. We added antibodies sequentially because anti-LacZ and anti-MH27 antibodies were from the same species. All conjugated secondary antibodies were from Jackson ImmunoResearch. Incubations overnight with antibodies were conducted at 4 °C and unbound antibody was washed off with several changes of buffer for 4 h at room temperature. Fixed animals were first incubated with monoclonal mouse anti-LacZ (Promega) diluted 1:500. After being washed, these animals were incubated with Cy3-conjugated goat anti-mouse secondary antibody, diluted 1:300. Worms were again washed and then incubated with polyclonal rabbit anti-GFP (Molecular Probes), diluted 1:300, and monoclonal antibody MH27 (ref. 28), diluted 1:1000. After another round of washes, animals were incubated with Cy2-conjugated goat anti-rabbit and Cy5-conjugated goat-anti mouse secondary antibodies, both diluted 1:300. Worms were washed as before and 4,6-diamidino-2-phenylindole (Sigma) was added to the last wash at a final concentration of 0.5 µg ml⁻¹. Stained worms were kept in SlowFade reagent (Molecular Probes). For visualization, stained worms (5 µl) were mounted on a 2% agarose pad and viewed with a Nikon Eclipse E800 microscope. Photomicrographs were acquired with a Bio-Rad MRC 1024ES laser scanning confocal attachment.

Received 24 July; accepted 11 October 2002; doi:10.1038/nature01234.

- Greenwald, I. *Caenorhabditis elegans II* (eds Riddle, D. L., Blumenthal, T., Meyer, B. J. & Priess, J. R.) 519–541 (Cold Spring Harbor Press, Cold Spring Harbor, New York, 1997).
- Levitán, D. & Greenwald, I. LIN-12 protein expression and localization during vulval development in *C. elegans*. *Development* **125**, 3101–3109 (1998).

- Stevens, J. L., Cantin, G. T., Wang, G., Shevchenko, A. & Berk, A. J. Transcription control by E1A and MAP kinase pathway via Sur2 mediator subunit. *Science* **296**, 755–758 (2002).
- Singh, N. & Han, M. *sur-2*, a novel gene, functions late in the *let-60* ras-mediated signaling pathway during *Caenorhabditis elegans* vulval induction. *Genes Dev.* **9**, 2251–2265 (1995).
- Wilkinson, H. A. & Greenwald, I. Spatial and temporal patterns of *lin-12* expression during *C. elegans* hermaphrodite development. *Genetics* **141**, 513–526 (1995).
- Wilkinson, H. A., Fitzgerald, K. & Greenwald, I. Reciprocal changes in expression of the receptor *lin-12* and its ligand *lag-2* prior to commitment in a *C. elegans* cell fate decision. *Cell* **79**, 1187–1198 (1994).
- Burdine, R. D., Branda, C. S. & Stern, M. J. EGL-17 (FGF) expression coordinates the attraction of the migrating sex myoblasts with vulval induction in *C. elegans*. *Development* **125**, 1083–1093 (1998).
- Grant, B. & Greenwald, I. Structure, function, and expression of SEL-1, a negative regulator of LIN-12 and GLP-1 in *C. elegans*. *Development* **124**, 637–644 (1997).
- Fire, A., Harrison, S. W. & Dixon, D. A modular set of lacZ fusion vectors for studying gene expression in *Caenorhabditis elegans*. *Gene* **93**, 189–198 (1990).
- Sandoval, I. V., Martinez-Arca, S., Valdeuza, J., Palacios, S. & Holman, G. D. Distinct reading of different structural determinants modulates the dileucine-mediated transport steps of the lysosomal membrane protein LIMP2 and the insulin-sensitive glucose transporter GLUT4. *J. Biol. Chem.* **275**, 39874–39885 (2000).
- Kirchhausen, T. Adaptors for clathrin-mediated traffic. *Annu. Rev. Cell Dev. Biol.* **15**, 705–732 (1999).
- Dietrich, J. *et al.* Molecular characterization of the di-leucine-based internalization motif of the T cell receptor. *J. Biol. Chem.* **271**, 11441–11448 (1996).
- Dittrich, E., Haft, C. R., Muys, L., Heinrich, P. C. & Graeve, L. A di-leucine motif and an upstream serine in the interleukin-6 (IL-6) signal transducer gp130 mediate ligand-induced endocytosis and down-regulation of the IL-6 receptor. *J. Biol. Chem.* **271**, 5487–5494 (1996).
- Ehrlich, M., Shmueli, A. & Henis, Y. I. A single internalization signal from the di-leucine family is critical for constitutive endocytosis of the type II TGF-beta receptor. *J. Cell Sci.* **114**, 1777–1786 (2001).
- Kil, S. J. & Carlin, C. EGF receptor residues leu⁶⁷⁹, leu⁶⁸⁰ mediate selective sorting of ligand-receptor complexes in early endosomal compartments. *J. Cell. Physiol.* **185**, 47–60 (2000).
- Fitzgerald, K., Wilkinson, H. A. & Greenwald, I. *glp-1* can substitute for *lin-12* in specifying cell fate decisions in *Caenorhabditis elegans*. *Development* **119**, 1019–1027 (1993).
- Grant, B. & Hirsh, D. Receptor-mediated endocytosis in the *Caenorhabditis elegans* oocyte. *Mol. Biol. Cell* **10**, 4311–4326 (1999).
- Fares, H. & Greenwald, I. Genetic analysis of endocytosis in *Caenorhabditis elegans*: coelomocyte uptake defective mutants. *Genetics* **159**, 133–145 (2001).
- Nonet, M. L. Visualization of synaptic specializations in live *C. elegans* with synaptic vesicle protein-GFP fusions. *J. Neurosci. Methods* **89**, 33–40 (1999).
- Hubbard, E. J., Wu, G., Kitajewski, J. & Greenwald, I. *sel-10*, a negative regulator of *lin-12* activity in *Caenorhabditis elegans*, encodes a member of the CDC4 family of proteins. *Genes Dev.* **11**, 3182–3193 (1997).
- Heitzler, P. & Simpson, P. Altered epidermal growth factor-like sequences provide evidence for a role of Notch as a receptor in cell fate decisions. *Development* **117**, 1113–1123 (1993).
- Jacobsen, T. L., Brennan, K., Arias, A. M. & Muskavitch, M. A. Cis-interactions between Delta and Notch modulate neurogenic signalling in *Drosophila*. *Development* **125**, 4531–4540 (1998).
- Chitnis, A., Henrique, D., Lewis, J., Ish-Horowitz, D. & Kintner, C. Primary neurogenesis in *Xenopus* embryos regulated by a homologue of the *Drosophila* neurogenic gene Delta. *Nature* **375**, 761–766 (1995).
- Heitzler, P., Bourouis, M., Ruel, L., Carteret, C. & Simpson, P. Genes of the Enhancer of split and achaete-scute complexes are required for a regulatory loop between Notch and Delta during lateral signalling in *Drosophila*. *Development* **122**, 161–171 (1996).
- Dubois, L., Lecourtis, M., Alexandre, C., Hirst, E. & Vincent, J. P. Regulated endocytic routing modulates wingless signaling in *Drosophila* embryos. *Cell* **105**, 613–624 (2001).
- Mello, C. & Fire, A. *Caenorhabditis elegans: Modern Biological Analysis of an Organism* (eds Epstein, H. F. & Shakes, D. C.) 451–482 (Academic, San Diego, 1995).
- Miller, D. M. III & Niemeyer, C. J. Expression of the *unc-4* homeoprotein in *Caenorhabditis elegans* motor neurons specifies presynaptic input. *Development* **121**, 2877–2886 (1995).
- Priess, J. R. & Hirsh, D. I. *Caenorhabditis elegans* morphogenesis: the role of the cytoskeleton in elongation of the embryo. *Dev. Biol.* **117**, 156–173 (1986).
- Ambros, V. Cell cycle-dependent sequencing of cell fate decisions in *Caenorhabditis elegans* vulva precursor cells. *Development* **126**, 1947–1956 (1999).
- Tamura, K. *et al.* Physical interaction between a novel domain of the receptor Notch and the transcription factor RBP-J kappa/Su(H). *Curr. Biol.* **5**, 1416–1423 (1995).

Supplementary Information accompanies the paper on Nature's website (http://www.nature.com/nature).

Acknowledgements We thank M. Stern for the *egl-17p::lacZ* plasmid and helpful advice on making the modified *egl-17* promoter; R. Waterston for anti-MH27 antibody; M. Han for the *sur-2(ku9)* strain; H. Fares for reagents and much helpful advice; K. Simpson for help with backcrossing integrated arrays; B. Grant for anti-sera against endocytic markers; D. Levitan for advice on immunostaining; R. Ruiz for technical assistance; and B. Grant, O. Hobert, S. Jarriault, X. Karp, G. Struhl and A. Tomlinson for discussion and critical reading of the manuscript. D.S. was supported by a National Science Foundation predoctoral fellowship and an NIH National Institute on Aging training grant, coordinated by R. Liem. I.G. is an Investigator of the Howard Hughes Medical Institute.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to I.G. (e-mail: greenwald@cancercenter.columbia.edu).

Signal-dependent regulation of splicing via phosphorylation of Sam68

Nathalie Matter*†, Peter Herrlich*‡ & Harald König*

* Forschungszentrum Karlsruhe GmbH, Institut für Toxikologie und Genetik, and
‡ Universität Karlsruhe, Institut für Genetik, Postfach 3640, D-76021 Karlsruhe, Germany

Evolution of human organismal complexity from a relatively small number of genes^{1,2}—only approximately twice that of worm or fly—is explained mainly by mechanisms generating multiple proteins from a single gene, the most prevalent of which is alternative pre-messenger-RNA splicing^{1,3,4}. Appropriate spatial and temporal generation of splice variants demands that alternative splicing be subject to extensive regulation, similar to transcriptional control. Activation by extracellular cues of several cellular signalling pathways can indeed regulate alternative splicing^{5–8}. Here we address the link between signal transduction and splice regulation. We show that the nuclear RNA-binding protein Sam68 is a new extracellular signal-regulated kinase (ERK) target. It binds exonic splice-regulatory elements of an alternatively spliced exon that is physiologically regulated by the Ras signalling pathway, namely exon v5 of CD44. Forced expression of Sam68 enhanced ERK-mediated inclusion of the v5-exon sequence in mRNA. This enhancement was impaired by mutation of ERK-phosphorylation sites in Sam68, whereas ERK phosphorylation of Sam68 stimulated splicing of the v5 exon *in vitro*. Finally, Ras-pathway-induced alternative splicing of the endogenous CD44-v5 exon was abolished by suppression of Sam68 expression. Our data define Sam68 as a prototype regulator of alternative splicing whose function depends on protein modification in response to extracellular cues.

Alternatively spliced isoforms of CD44 are generated by the inclusion of up to 10 variant exon sequences (v1–v10) in CD44 mRNA and are decisive in embryonic development, in immune response and in tumour progression^{9–12}. We previously identified RNA elements necessary for signalling-induced alternative splicing in CD44 exon v5 (ref. 5), an exon frequently included in the course of tumour progression and upon T-cell activation^{8,12}, and we showed that in T lymphocytes and lymphoma cells alternative splicing is regulated through these elements by the ERK mitogen-activated protein (MAP) kinase pathway⁸, an evolutionarily conserved enzyme cascade connecting cell surface receptors to regulatory targets¹³. Recently, Stoss *et al.* in collaboration with us linked Etoile/Sam68-like mammalian protein (SLM-2) to alternative splicing¹⁴. SLM-2 belongs to the STAR (signal transduction and activation of RNA) protein family¹⁵, whose domain organization suggests RNA binding and modification in response to signal transduction. This led to the hypothesis that STAR proteins could be involved in coupling signal transduction to alternative pre-mRNA splicing¹⁴.

To test this hypothesis, we first measured the binding of STAR proteins to the splice regulatory elements in the CD44 v5 exon. Because SLM-2 is not expressed in our T-lymphoma cells used to study alternative splicing *in vivo* (H.K., unpublished observations), we screened these cells for expression of other family members and found an abundant presence of Sam68 (Src-associated in mitosis)^{16,17}. Endogenous (Fig. 1a, lanes 2–4) or recombinant glutathione S-transferase (GST)-tagged Sam68 (Fig. 1a, lanes 7–9) were

indeed affinity-precipitated with the exonic v5 splice-regulatory elements (v5 subdomains L, M and R; see the scheme in Fig. 1a). Barely any association with two unrelated control sequences⁵ was detectable (Fig. 1a, lanes 1, 5, 6 and 10). The GST portion alone did not interact (data not shown). Binding of GST–Sam68 to subdomain L, but not to subdomains M and R, could also be demonstrated in electrophoretic mobility-shift assays (EMSA) (Fig. 1b, lanes 1–6), indicating that only the interaction of Sam68 with subdomain L is stable enough to survive EMSA gel resolution. In keeping with an enhanced affinity to L, linker-scan mutants (data not shown) narrowed down Sam68 binding to 10 nucleotides in the L subdomain comprising the sequence AAAAUU (see Fig. 1b), which resembles high-affinity Sam68-binding sites previously defined by SELEX¹⁸. Replacing two nucleotides in this sequence by cytidine (L/CC) abolished both complex formation with Sam68 (Fig. 1b, lanes 7 and 8) and competition for Sam68 binding to the wild-type L subdomain (Fig. 1b, lanes 9–13), suggesting this sequence element to be a genuine Sam68-binding site. In conclusion, the RNA-binding data suggest that, although Sam68 binds subdomain L with higher affinity, it can interact with all three regulatory subdomains of the v5 exon by direct RNA binding.

To examine whether Sam68 can affect inclusion of the CD44 v5-exon sequence in mRNA, and whether it might do so in a signal-dependent manner, we co-transfected a splice-reporter minigene⁵

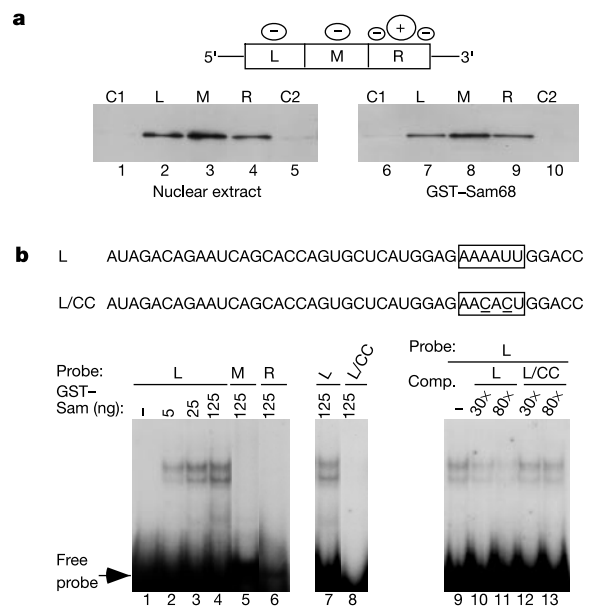


Figure 1 Sam68 binds splice-regulatory sequences in CD44 exon v5. **a**, RNA-affinity precipitations from EL4-cell nuclear extracts (lanes 1–5) and of recombinant GST–Sam68 (lanes 6–10) with biotinylated oligonucleotides corresponding to either subdomains L, M and R of CD44 exon v5 (see diagram) or to unrelated control sequences (C1, C2). After SDS–polyacrylamide gel electrophoresis (SDS–PAGE), proteins were subjected to immunoblot analysis with anti-Sam68 monoclonal antibody 7-1. Bands correspond to the expected relative molecular masses (M_r) of Sam68 and GST–Sam68: 68,000 (68K) and 90K, respectively. The diagram shows the domain organization of the v5 exon: L and M were defined as splice silencer regions (–) preventing exon inclusion in cell types or under conditions in which the exon is not used, whereas R, in addition to negative regulatory sequences (–), harbours a splice enhancer (+) necessary for exon inclusion⁵. **b**, EMSA using recombinant GST–Sam68 protein and ³²P-labelled RNA probes corresponding to L, M and R or to the mutant L/CC. In lanes 9–13, 25 ng of GST–Sam68 was used. Comp., non-labelled competitor oligonucleotides (30-fold or 80-fold excess over the labelled probe). Sequences of subdomain L and the L/CC mutant are shown; an element similar to Sam68 high-affinity binding sequences defined by SELEX¹⁸ is boxed; mutated nucleotides are underlined.

† Present address: Institut de Génétique et de Biologie Moléculaire et Cellulaire INSERM – U.184/CNRS – LGME/ULP, BP 10142, F-67404 Illkirch Cedex, C.U. de Strasbourg, France.

carrying CD44 exon v5 (see Fig. 2a) with an expression construct for murine Sam68 or the empty expression vector into mouse EL4 T-lymphoma cells, in which v5-exon inclusion can be stimulated by treatment with phorbol ester. Forced expression of Sam68 affected only weakly the basal level of v5-exon inclusion in analysis by

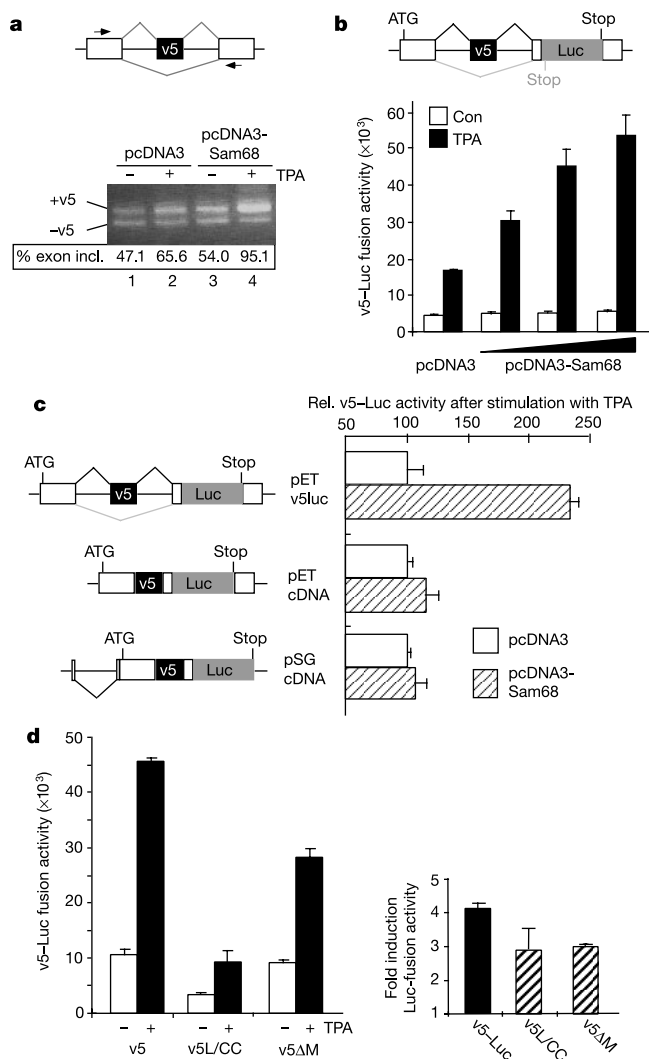


Figure 2 Sam68 regulates the inclusion of CD44 variant exon v5. **a**, RT-PCR analysis of EL4 cells transiently co-transfected with the CD44v5-minigene construct pETv5 (ref. 5) and either an expression construct for murine Sam68 (pcDNA3-Sam68) or the empty expression vector (pcDNA3). Cells were either treated with TPA (40 ng ml⁻¹) (+) or with DMSO (solvent control) (-) for 6 h before RNA preparation. Values for percentage of exon inclusion (% exon incl.) are boxed. Diagram shows the minigene with insulin exons (open boxes), introns (black lines), CD44 exon v5 (black box), splicing alternatives (grey lines) and PCR primer positions (arrows). **b**, v5-Luciferase (v5-Luc) fusion activity from EL4 cells co-transfected with the splice-reporter gene pETv5luc (named pETcatEBlucv5 in ref. 8) and either pcDNA3 (2.5 µg) or increasing amounts of pcDNA3-Sam68 (0.5, 1.5 and 2.5 µg). Treatment with TPA was as in **a**. The diagram shows the splice-reporter gene⁸ with insulin exon sequences (open boxes), CD44 exon v5 (black box), luciferase (Luc) coding sequence (grey box) and introns (black lines); v5-exon skipping (grey) and inclusion (black) modes are indicated. **c**, Relative v5-Luc activities after TPA stimulation from the reporter constructs depicted at the left, co-transfected with either pcDNA3 or pcDNA3-Sam68. Symbols as in **b** with β-globin-intron splicing indicated in pSGcDNA. **d**, v5-Luc fusion activity from EL4 cells transiently transfected with 5 µg of either the wild-type (v5-Luc), the L/CC-mutant (v5L/CC) or the M-deletion mutant (v5ΔM) CD44v5-luciferase splice-reporter constructs. Fold induction of luciferase activity after stimulation with TPA is shown at the right. Error bars indicate standard deviations based on three independent transfections.

polymerase chain reaction with reverse transcriptase (RT-PCR) (Fig. 2a, lanes 1 and 3); however, it markedly enhanced phorbol-ester-stimulated inclusion of the exon (lanes 2 and 4). For better quantification we performed similar co-transfection experiments with a luciferase-based version of the splice-reporter minigene (see Fig. 2b), which translates changes in v5-exon inclusion into changes in luciferase expression⁸. Again, Sam68 expression had no significant effect on luciferase splice-reporter activity in the absence of phorbol ester but it enhanced phorbol-ester-induced reporter activity in a dose-dependent manner (Fig. 2b), suggesting a change in factor activity. Because Sam68 has been implicated in nuclear export of retroviral RNA¹⁹, we wished to assess the induced nuclear export of the v5-containing mRNA as a putative cause of the effects observed. We therefore expressed v5-luciferase fusion mRNA either from an intronless version of the v5-splice reporter construct pETv5luc used in Fig. 2b (pETcDNA; see Fig. 2c) or from a vector containing the fusion complementary DNA preceded by β-globin intron II for constitutive splicing of the expressed v5 fusion RNA (pSGcDNA, see Fig. 2c), taking into account a possible splicing-dependent recruitment of mRNA-export factors. In marked contrast to the v5-splice reporter construct (pETv5luc), neither of these constructs showed a significant effect of Sam68 on stimulation by phorbol ester (Fig. 2c), making induced mRNA export highly unlikely as an underlying mechanism. Thus, the co-transfection experiments indicate that Sam68 abundance in EL4 cells is sub-optimal and strongly suggest that Sam68 enhances phorbol-ester-stimulated inclusion of CD44 exon v5.

Binding of Sam68 to v5-exon elements indicated a putative role for these sequences in mediating Sam68 activity. In support of this, we observed a 4–5-fold decrease in luciferase splice-reporter activity (Fig. 2d), reflecting decreased levels of v5-exon inclusion in RT-PCR analysis (data not shown) after introducing the L/CC mutation into the v5 exon. However, induction by phorbol ester was only mildly affected in the mutant, which might be attributed to additional Sam68-binding sites in subdomains M and R (as suggested in Fig. 1a). Consistent with this assumption, deletion of subdomain M (subdomain R cannot be removed without abolishing exon inclusion⁵) also impaired the induction of v5 reporter activity by phorbol ester (Fig. 2d). These results suggest that the binding of Sam68 to the high-affinity site in subdomain L is critical for the general level of v5-exon inclusion, whereas induction by phorbol ester might depend on multiple v5 exon sequences. In the co-transfection experiments (Fig. 2b), an effect of Sam68 expression on basal exon inclusion was probably undetectable owing to saturation of the high-affinity site in subdomain L by endogenous protein.

The response to phorbol ester raised the question of whether Sam68 was subjected to activation through signal transduction. Phorbol esters are known to activate the Ras-signalling pathway in T lymphocytes²⁰, and oncogenically activated Ras induced the inclusion of CD44 exon v5 dependent on ERK MAP-kinase signalling in T-lymphoma cells⁸. We therefore tested whether Sam68 action depended on ERK-pathway activation. U0126, a specific inhibitor of MAP-kinase ERK kinase²¹ (MEK1/2), the kinase activating ERK, almost completely abolished the phorbol-ester-induced enhancement of splice-reporter activity by co-transfected Sam68 (Fig. 3a). The activation of ERK MAP-kinase therefore seems crucial for Sam68 activity. Given that phorbol-ester-stimulated v5-exon inclusion occurs independently of protein synthesis *de novo*⁵, we examined whether Sam68 could be a target for regulatory post-translational modification. Electrophoretic mobility of Sam68 revealed a partial shift in apparent molecular mass after cells had been treated with phorbol ester (Fig. 3b, top panel, lanes 1 and 2; and Fig. 3c, top panel). This shift was prevented by the MEK inhibitor U0126 (Fig. 3b, top panel, lanes 3 and 4). We conclude that a subfraction of Sam68 molecules was modified, most probably phosphorylated, in a manner requiring ERK activation.

Like unmodified Sam68, the modified subfraction was completely nuclear (data not shown).

The murine Sam68 sequence shows eight potential proline-directed ERK phosphorylation sites¹³, six of which comprise threonine followed by proline (TP). To determine their putative phosphorylation, we employed a phosphothreonine-proline (phospho-TP)-specific antibody and detected TP phosphorylation of the shifted Sam68 fraction (Fig. 3b, second panel, lanes 1 and 2). Like the shift in apparent molecular mass, TP phosphorylation was prevented by treatment with U0126 (lanes 3 and 4). It coincided with phosphorylation (indicating activation) of ERK MAP-kinase (Fig. 3b, bottom two panels) and occurred as early as 1 minute after treatment with phorbol ester (Fig. 3c). According to the time course of induced phosphorylation, Sam68 could be a direct substrate for ERK MAP-kinase. In support of this assumption, purified GST-Sam68 protein was phosphorylated by recombinant ERK2 in a concentration-dependent manner (Fig. 3d) as well as by ERK1/2 immunoprecipitated from phorbol ester-stimulated EL4 cells (data not shown). Taken together, these data define Sam68 as a novel ERK target and strongly indicate that TP phosphorylation of Sam68

observed after activation of the ERK MAP-kinase pathway *in vivo* is mediated directly by ERK MAP-kinase.

If ERK phosphorylation of Sam68 were indeed involved in Sam68 activation, replacement by alanine of all eight threonine and serine residues of the putative MAP-kinase target sites in Sam68 should impair the induction of its splice-activating potential. When

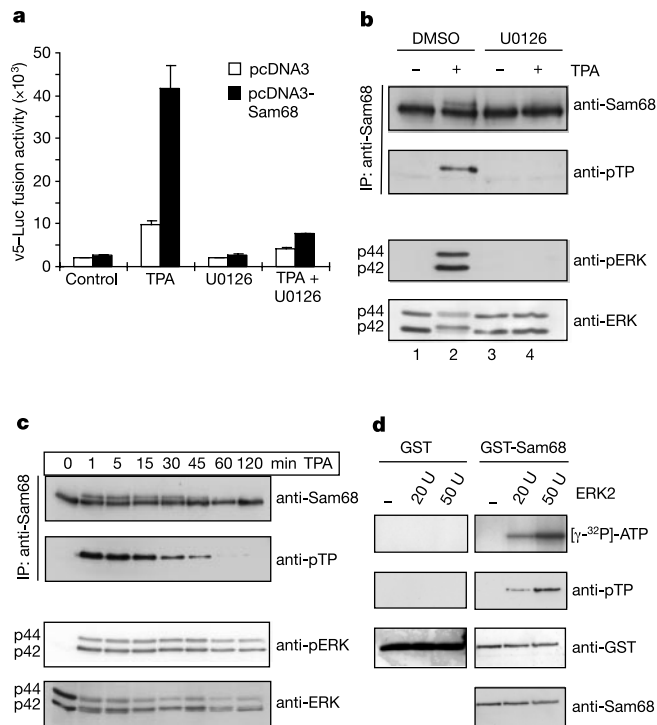


Figure 3 Activation and phosphorylation of Sam68 by ERK. **a**, TPA-induced v5-luciferase splice-reporter activity of EL4 cells co-transfected with pcDNA3-Sam68 or pcDNA3, in the absence or the presence of the MEK inhibitor U0126 (20 μ M; Promega). Treatment with TPA was as in Fig. 2a. Error bars indicate standard deviations based on three independent transfections. **b**, Western blot analysis (as indicated at the right) of immunoprecipitated (IP) Sam68 and of lysates from EL4 cells that were either left untreated (–) or treated with 40 ng ml^{–1} TPA (+) in the absence or the presence of U0126. Immunoblot analysis of Sam68 was performed with anti-Sam68 antibody C20 and an antibody directed against phosphothreonine-proline (anti-pTP), respectively. The non-modified Sam68 band shows the expected M_r value of 68K. Analysis of ERK was performed by phospho-p44/p42 MAP kinase (Thr 202/Tyr 204) and p44/p42 MAP kinase polyclonal antibodies (Cell Signaling Technology), respectively. **c**, Time course of Sam68 and ERK MAP-kinase phosphorylation after treatment of EL4 cells with TPA (40 ng ml^{–1}). Analysis was as in **b**. **d**, Phosphorylation of GST-Sam68 fusion protein *in vitro* without (–) or with 20 or 50 units (U) of recombinant activated p42 MAP kinase (ERK2). GST was used as a control. After SDS-PAGE, products were either subjected to autoradiography ([γ -³²P]ATP) or analysed by immunoblotting with the antibodies indicated at the right. Bands correspond to M_r values of 26K for GST and 90K for GST-Sam68.

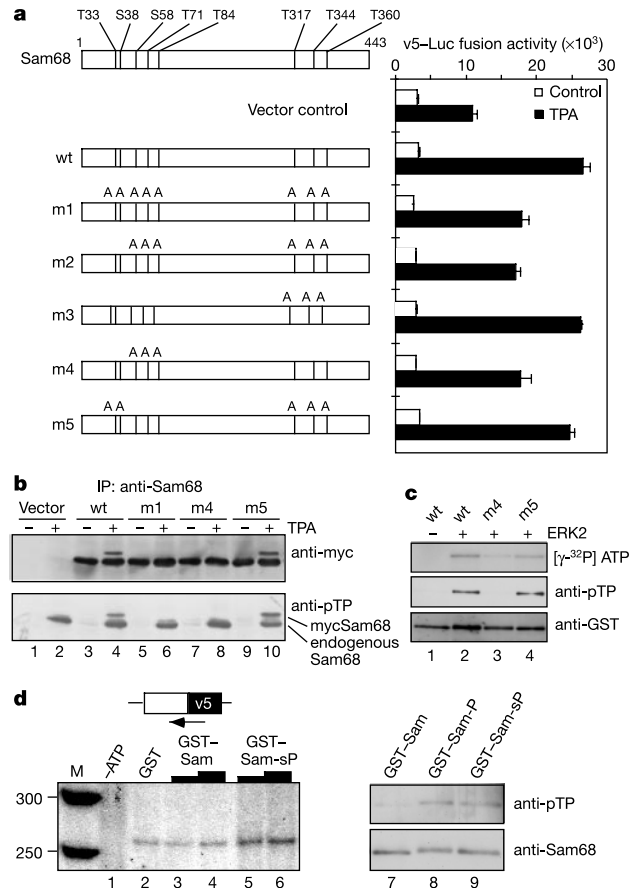


Figure 4 Phosphorylation of Sam68 at ERK-target sites stimulates v5-exon usage *in vivo* and *in vitro*. **a**, Enhancement of v5-splice-reporter activity by phorbol ester in EL4 cells co-transfected with pcDNA3 expression constructs for wild-type Sam68 (wt) or for the phosphorylation-site mutants indicated (m1–m5). Diagram shows serine (S) and threonine (T) residues of Sam68 replaced by alanine (A). Phorbol ester (TPA) and control treatments were as in Fig. 2a. Error bars indicate standard deviations based on three independent transfections. **b**, Western blot analysis (with antibodies indicated at the right) of Sam68 immunoprecipitations (IP) from EL4 cells transiently transfected with either the empty expression vector or expression vectors for Myc-epitope-tagged wild-type (wt) or mutant (m1, m4 and m5) Sam68 proteins. DMSO (solvent control) (–) or 40 ng ml^{–1} TPA (+) were added 5 min before lysate preparation. Western blotting was performed with anti-c-Myc antibody 9E10 (Santa Cruz Biotechnology) and an anti-phosphothreonine-proline antibody (anti-pTP), respectively. Phosphorylated mycSam68 and endogenous Sam68 proteins are indicated. **c**, *In vitro* phosphorylation of wild-type GST-Sam68 (wt) and of the GST-Sam68 mutants indicated, with the use of recombinant activated ERK2. Reactions were performed in the absence (–) or presence (+) of 100 units of ERK2 and analysed as in Fig. 3d. **d**, Lanes 1–6: *in vitro* splicing reactions in the presence of GST (3 ng μ l^{–1}), mock-phosphorylated (GST-Sam) or thiophosphorylated (GST-Sam-sP) GST-Sam68 (1.5 and 3 ng μ l^{–1} each). ‘– ATP’ indicates GST control reaction without ATP. Exon-v5 splicing to insulin exon 2 was detected in pre-mRNA generated from a derivative of the pETv5 splice reporter⁵ (used in Fig. 2a) by primer-extension, using a splice-junction primer (see Methods). The expected size of the extension product is 272 nucleotides. M, molecular mass marker (sizes in nucleotides). The scheme shows spliced RNA with insulin exon 2 (open box), CD44 exon v5 (black box) and primer (arrow). Lanes 7–9: Western blot analysis of GST-Sam68 proteins used in the *in vitro* splicing reactions. GST-Sam, mock-phosphorylated protein; GST-Sam-P and GST-Sam-sP, proteins phosphorylated in the presence of ATP and ATP- γ S, respectively.

co-transfected with the CD44-v5 splice-reporter construct, the eight-amino-acid exchange mutant was significantly impaired in its ability to mediate phorbol-ester-induced exon inclusion (Fig. 4a, mutant m1). By swapping Sam68 portions comprising mutated and non-mutated ERK target sites, we found that Sam68 mutated in Ser 58, Thr 71 and Thr 84 showed the same extent of impairment in induced exon inclusion as did Sam68 mutated in all S/TP sites (Fig. 4a, mutant m4). Conversely, mutants in which these residues were unchanged behaved like the wild-type Sam68 protein (Fig. 4a, mutants m3 and m5). Single mutations of residues Ser 58, Thr 71 or Thr 84 showed no significant effects. Binding of RNA by multiple phosphorylation-site mutants was comparable to that of the wild-type protein (data not shown), excluding reduced binding of RNA as a cause of their impaired function. Myc-epitope-tagged versions of the total S/TP → A mutant (m1) as well as of the the S58A-T71A-T84A-triple mutant (m4) expressed in EL4 cells no longer produced the phorbol-ester-induced shift in apparent molecular mass (Fig. 4b, upper panel, lanes 3–8) nor the additional (mycSam68) TP-phosphorylated band (lower panel, lanes 3–8), whereas the complementary mutant (m5) did (lanes 9 and 10). In agreement with this result, incorporation of $^{32}\text{PO}_4$ into recombinant GST–Sam68 proteins by incubation with ERK2 and [$\gamma\text{-}^{32}\text{P}$]ATP *in vitro* was severely reduced, and overall TP phosphorylation was not detectable when using the triple mutant (m4) (Fig. 4c, lanes 2 and 3). Incorporation of $^{32}\text{PO}_4$ into m5 mutant protein was affected to a smaller extent, and phospho-TP antibody staining was comparable to that of wild-type GST–Sam68 (Fig. 4c, lanes 2 and 4). Thus, Ser 58, Thr 71 and Thr 84 are the predominant ERK-phosphorylation sites. Proline-directed phosphorylation of these sites is, at least in part, required for induced alternative

splicing of exon v5 after activation of the ERK MAP-kinase pathway. Residual activity of the phosphorylation-site mutants after treatment with phorbol ester could be explained by the recruitment of ERK through the mutants into complexes containing endogenous Sam68. Alternatively, additional ERK targets recruited by Sam68 or phosphorylated dependent on ERK recruitment through Sam68 might participate in splice regulation.

To test whether Sam68 and its phosphorylation by ERK can affect v5 splicing directly, we set up *in vitro* splicing reactions supplemented with either non-phosphorylated or ERK-phosphorylated recombinant GST–Sam68 protein. Because the phosphorylation of Sam68 by ERK is very rapidly reverted in nuclear extracts (data not shown) and splicing is much less efficient *in vitro* than *in vivo* (making long incubation times necessary), we reasoned that Sam68 phosphorylation must be stabilized to see putative splice activation *in vitro*. Because phosphatase inhibitors block the catalytic steps of splicing²², we used ATP- γS to stably phosphorylate Sam68 by ERK. Thiophosphorylation of GST–Sam68 led to a similar shift in apparent molecular mass and reactivity to phospho-TP-specific antibody as observed with the ATP-phosphorylated protein (Fig. 4d, lanes 7–9). It prevented dephosphorylation but did not change protein stability in nuclear extracts (data not shown). The extensive pre-mRNA length and low efficiency of v5 splicing led us to detect spliced transcripts by primer extension with a splice-junction primer (see the scheme in Fig. 4d) specifically recognizing the spliced RNA (Fig. 4d, compare lanes 1 and 2). In contrast to mock-phosphorylated GST–Sam68 preparations, thiophosphorylated GST–Sam68 markedly stimulated v5 splicing compared with the GST control (Fig. 4d, lanes 2–6), suggesting direct splice regulation by Sam68 phosphorylation.

Thus, the data indicate that Sam68 can act as splice factor for CD44-v5 alternative splicing whose activity is regulated by ERK phosphorylation. To investigate whether Sam68 is crucial to splice regulation of endogenous CD44 by Ras-pathway activation, we downregulated Sam68 expression in EL4 cells by morpholino antisense technology^{23,24}. Sam68-protein levels were specifically decreased by more than 80% (as quantified by western blotting) with the cognate Sam68-morpholino sequence, but not when a nonspecific control sequence was used (Fig. 5a). When analysing the expression of exon-v5-containing endogenous CD44 mRNA, the phorbol-ester-dependent induction observed in the control morpholino-treated cells (Fig. 5b, lanes 1 and 2) was almost undetectable in the Sam68-morpholino 'knockdown' cells (lanes 3 and 4). Basal v5 expression was not affected, possibly owing to occupancy of the high-affinity binding site in the L portion of the exon by residual Sam68 protein. Similarly to murine primary T cells and other T-lymphoma cells^{5,8}, treatment of EL4 cells with phorbol ester stimulated the expression of CD44 mRNAs containing primarily single variant exons, of which v1, v3, v5 and v7 are most reliably detectable. Interestingly, induction of v1, v3 and v7 expression by phorbol ester was not affected by 'knocking down' Sam68 expression (Fig. 5b). Thus, the data suggest exon specificity of Sam68 action and indicate a crucial role for Sam68 in Ras-pathway-induced alternative splicing of exon v5 in endogenous CD44.

The results define a first splice regulator protein that is linked to signal transduction and modulated directly by ERK and is therefore under the control of Ras and growth-factor signalling. We postulate that splice regulation will become just as complex as transcriptional control to account for the temporal and spatial requirements of gene expression, which needs to exploit a limited number of genes. The novel ERK substrate, Sam68, acts exon-specifically and does not seem to change affinity to RNA after phosphorylation *in vitro*. It might act as regulatory component in the splice-silencer complex formed over the v5-exon sequence^{5,25} by controlling protein-protein contacts to factors involved in splice-site recognition and/or in spliceosome assembly. Our data now open the subject of how

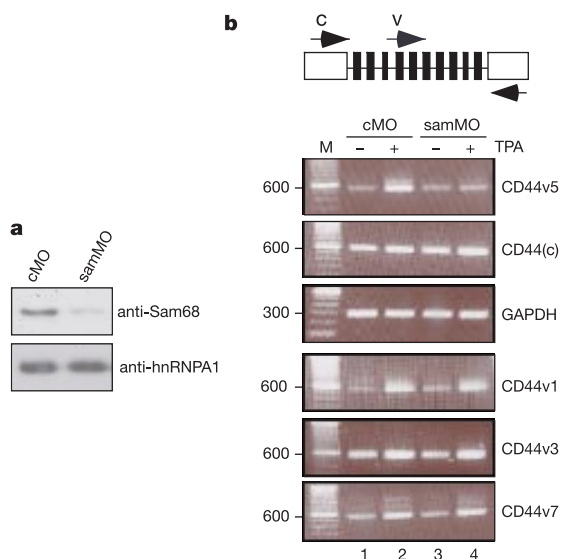


Figure 5 Sam68 is required for phorbol-ester regulation of v5-exon usage in endogenous CD44. **a**, Downregulation of Sam68 expression by antisense Sam68-morpholino sequence. A cognate Sam68-morpholino oligonucleotide (samMO) or a nonspecific control-morpholino sequence (cMO) were delivered to EL4 cells and Sam68 protein levels were analysed by western blotting. The blot was re-probed with hnRNP A1 antibody to check for protein loading. Bands correspond to M, values of 68K for Sam68 and 34K for hnRNP A1. **b**, RT-PCR analysis of cells in **a** that were either mock-treated (–) or treated with TPA (40 ng ml^{–1}) for 6 h. Diagram shows CD44 gene organization with constant-exon regions (open boxes), variant exons (closed boxes) and primer positions (arrows). Exon-specific PCRs were performed with 5' primers⁵ specific for the CD44 variant (v) exons indicated and a common 3' primer⁵ (see the diagram). The additional band in v5 PCR (lane 2) represents CD44v5/v6 and v5/v7 mRNA as analysed by sequencing. Total CD44, using a 5' constant region primer⁵ (c), and glyceraldehyde3phosphate dehydrogenase mRNA levels were monitored to control for CD44 transcriptional activation and cDNA amounts, respectively. M, 100-bp ladder (indicated sizes in base pairs).

exon choice is controlled by the signalling-induced modification of splice factors in normal physiology and disease. □

Methods

Cell culture and transfections

We cultured EL4 cells in DMEM with 10% fetal calf serum, 100 U ml⁻¹ penicillin and 100 µg ml⁻¹ streptomycin. Co-transfections (10⁷ cells) were performed with DEAE-Dextran, using 1 µg of reporter plasmid and the indicated amounts of Sam68 expression plasmid or empty vector. For analysis of mycSam68 modification, 2.5 × 10⁷ cells were transfected with 20 µg of plasmid DNA by electroporation.

Plasmid construction

pcDNA3-Sam68 was generated by cloning the murine Sam68 cDNA, amplified from the plasmid pcDNA3-mycSam68wt using *Pwo* polymerase and the primers 5'-CGGGATCCCGATGCAGCGCCGGGACGATCCTGCTCG-3' and 5'-CGGAATTCTTAATAACGTCATATGGATGCTCTC-3', into the *Bam*HI and *Eco*RI sites of pcDNA3.1 + (Invitrogen). Corresponding expression constructs for phosphorylation-site mutants were generated by replacing serines and threonines with alanine through site-directed mutagenesis (Quickchange kit; Stratagene). pETv5L/Ccluc and pETv5ΔMluc were generated by site-directed mutagenesis of pETv5luc and *Mlu*I deletion of pETCatEblucv5ΔM⁸, respectively. To obtain pETcDNA and pSGcDNA, insulin-v5-luciferase cDNA was cloned into the *Hind*III and *Bgl*II sites of pETCatEB⁸ and of a modified version of pSG5 (Stratagene), respectively. pSKv5ΔAvr/Sp resulted from insertion of an 1.2-kilobase *Pvu*II/*Eco*NI fragment of pETv5 (ref. 5) containing insulin exon2, CD44 exon v5 and the intervening intron into the *Sma*I site of pBluescript SK II (Stratagene), followed by a reduction of the intron size to 248 base pairs (bp) by *Avr*II/*Spe*I-fragment deletion.

RNA-binding assays

RNA-affinity precipitations were performed as described²⁵ by using 40 µg of nuclear extract and 0.8 µg of recombinant GST-Sam68 protein, respectively. Oligonucleotide sequences for L, M, R and C1 (named 'blue' in ref. 25) are given elsewhere²⁵; C2, 5'-CGACGCGUAUUAUGGUUAUGG CAGCACUGCAUACGCGUCG-3'. EMSA reactions were performed in 25 µl of binding buffer (10 mM HEPES pH 7.9, 100 mM KCl, 0.025% Nonidet P40, 10% glycerol, 1 mM dithiothreitol (DTT), 1 mM phenylmethanesulphonyl fluoride (PMSF)) in the presence of 3 µg of yeast tRNA and 0.05 pmol of labelled RNA oligonucleotide (20,000 c.p.m.) at room temperature for 15 min. Reactions were resolved on 4% native polyacrylamide gels in 0.5 × TBE. Probe sequences were as described²⁵, lacking biotinylation.

Cell lysate preparation and immunoprecipitation

After being washed once with PBS, cells (2 × 10⁷) were lysed in 800 µl of RIPA buffer (50 mM Tris-HCl pH 8.0, 125 mM NaCl, 0.5% IGEPAL CA-630 (Sigma), 0.5% sodium deoxycholate, 0.1% SDS) supplemented with 1 mM DTT, 1 mM PMSF, 2 µg ml⁻¹ aprotinin, 2 µg ml⁻¹ leupeptin, 2 µg ml⁻¹ pepstatin, 10 mM β-glycerolphosphate, 10 mM NaF, 300 mM okadaic acid, 50 µM dephostatin, 10 µM cypermethrin, 2 mM Na₃VO₄ and 50 mM calyculin A. Sam68 was immunoprecipitated by standard procedures with anti-Sam68 polyclonal antibody C20 (Santa Cruz Biotechnology) and Protein G Plus-agarose (Calbiochem).

In vitro phosphorylation assay

Approximately 0.8 µg of bacterially expressed GST-(mouse)Sam68 fusion proteins, or the same amount of GST as a control, were mixed in 40-µl reactions with 20–100 units of activated recombinant ERK2 (Cell Signaling Technology) in 1 × MAPK buffer (Cell Signaling Technology), in the presence of 100 µM ATP. Reactions were split, one half was supplemented with 5 µCi of [γ-³²P]ATP and reactions were incubated at 30 °C for 30 min. After resolution by SDS-PAGE, [γ-³²P]ATP-labelled reactions were subjected to autoradiography, non-radioactive reactions were analysed by immunoblotting with antibodies against phosphothreonine-proline (Cell Signaling Technology), against GST (2D5) or against Sam68 (clone 7-1; Santa Cruz Biotechnology).

GST fusion proteins

GST-Sam68 and GST-Sam68 mutants were expressed in *Escherichia coli* BL21 from constructs carrying corresponding Sam68 complementary DNAs (starting with the second amino acid) in the *Eco*RI site of pGEX-3X (Amersham Pharmacia). Fusion proteins were purified by the GST-purification module (Amersham Pharmacia) in accordance with the instructions of the manufacturer. Purified proteins were dialysed against 20 mM Tris-HCl pH 7.5, 50 mM NaCl, 2 mM EDTA, 1 mM DTT, 50% glycerol.

In vitro splicing

Capped pre-mRNA (713 nucleotides; nt) was transcribed *in vitro* with T3 RNA polymerase from linearized pSKv5ΔAvr/Sp. Nuclear extracts were prepared by a modified version of that in Dignam *et al.*²⁶ Splicing reactions (18.5 µl) contained 4 fmol µl⁻¹ pre-mRNA, 32.4% HeLa nuclear extract (4.6 µg µl⁻¹), 1.1 mM ATP, 2.9 mM MgCl₂, 21.6 mM creatine phosphate, 0.9 U µl⁻¹ RNasin, 64.9 mM KCl, 13 mM HEPES pH 7.9, 0.13 mM EDTA, 13% glycerol, 0.65 mM DTT and 3.2% polyvinyl alcohol. After incubation at 30 °C for 90 min, reactions were deproteinized by proteinase K and extraction with phenol/chloroform. RNA was precipitated with ethanol and analysed by primer extension with a ³²P-labelled splice-junction primer complementary to the last nine nucleotides of insulin exon 2 and to the first 21 nucleotides of CD44 exon v5. Thiophosphorylation (in the presence of 2 mM ATP-γS) or mock-phosphorylation (in the absence of ATP) of recombinant Sam68 by

ERK was as described above, followed by dialysis against Dignam buffer D. Splicing reactions were performed with two different preparations of nuclear extracts and of recombinant Sam68 with entirely consistent results.

Morpholino antisense experiments

Pre-paired morpholino oligonucleotides were obtained from GeneTools, LCC. An antisense Sam68 sequence (samMO, 5'-GGGATGATGGGTGGCGAGCGAACGA-3') or a standard control sequence²⁴ (cMO, 5'-CCTCTTACCTCAGTTACAATTATA-3') was delivered by electroporation (10 nmol per 1.3 × 10⁷ cells). After 24 h, cells were either mock-treated (DMSO) or treated with 12-*O*-tetradecanoylphorbol-13-acetate (TPA) for 6 h; 4 × 10⁶ cells were used for the preparation of cell lysates (as described above) and for that of cytoplasmic RNA⁵. Western blot analysis of Sam68 was performed as described above; heterogeneous ribonucleoprotein A1 (hnRNP A1) was detected by monoclonal antibody 4B10. RT-PCR was performed essentially as described⁵ with 38 cycles for CD44 and 25 cycles for GAPDH amplifications. PCR amplifications were in linear phase as verified with different amounts of cDNA.

Received 25 June; accepted 6 September 2002; doi:10.1038/nature01153.

1. International human genome sequencing consortium Initial sequencing and analysis of the human genome. *Nature* **409**, 860–921 (2001).
2. Venter, J. C. *et al.* The sequence of the human genome. *Science* **291**, 1304–1351 (2001).
3. Sharp, P. A. Split genes and RNA splicing. *Cell* **77**, 805–815 (1994).
4. Maniatis, T. & Tasic, B. Alternative pre-mRNA splicing and proteome expansion in metazoans. *Nature* **418**, 236–243 (2002).
5. König, H., Ponta, H. & Herrlich, P. Coupling of signal transduction to alternative pre-mRNA splicing by a composite splice regulator. *EMBO J.* **17**, 2904–2913 (1998).
6. van der Hoven van Oordt, W. *et al.* The MKK3/6-p38-signaling cascade alters the subcellular distribution of hnRNP A1 and modulates alternative splicing regulation. *J. Cell Biol.* **149**, 307–316 (2000).
7. Xie, J. & Black, D. L. A CaMK IV responsive RNA element mediates depolarization-induced alternative splicing of ion channels. *Nature* **410**, 936–939 (2001).
8. Weg-Remers, S., Ponta, H., Herrlich, P. & König, H. Regulation of alternative pre-mRNA splicing by the ERK MAP-kinase pathway. *EMBO J.* **20**, 4194–4203 (2001).
9. Arch, R. *et al.* Participation in normal immune response of a splice variant of CD44 that encodes a metastasis-inducing domain. *Science* **257**, 682–685 (1992).
10. Cooper, D. L. & Dougherty, G. J. To metastasize or not? Selection of CD44 splice sites. *Nature Med.* **1**, 635–637 (1995).
11. Sherman, L., Wainwright, D., Ponta, H. & Herrlich, P. A splice variant of CD44 expressed in the apical ectodermal ridge presents fibroblast growth factors to limb mesenchyme and is required for limb outgrowth. *Genes Dev.* **12**, 1058–1071 (1998).
12. Sherman, L. *et al.* The CD44 proteins in embryonic development and in cancer. *Curr. Top. Microbiol. Immunol.* **213**, 249–269 (1996).
13. Chang, L. & Karin, M. Mammalian MAP kinase signalling cascades. *Nature* **410**, 37–40 (2001).
14. Stoss, O. *et al.* The STAR/GSG family protein rSLM-2 regulates the selection of alternative splice sites. *J. Biol. Chem.* **276**, 8665–8673 (2001).
15. Vernet, C. & Artzt, K. STAR, a gene family involved in signal transduction and activation of RNA. *Trends Genet.* **13**, 479–484 (1997).
16. Taylor, S. J. & Shalloway, D. An RNA-binding protein associated with Src through its SH2 and SH3 domains in mitosis. *Nature* **368**, 867–871 (1994).
17. Fumagalli, S., Totty, N. F., Hsuan, J. J. & Courtneidge, S. A. A target for Src in mitosis. *Nature* **368**, 871–874 (1994).
18. Lin, Q., Taylor, S. J. & Shalloway, D. Specificity and determinants of Sam68 RNA binding. *J. Biol. Chem.* **272**, 27274–27280 (1997).
19. Reddy, T. R., Tang, H., Xu, W. & Wong-Staal, F. Sam68, RNA helicase A and Tap cooperate in the post-transcriptional regulation of human immunodeficiency virus and type D retroviral mRNA. *Oncogene* **19**, 3570–3575 (2000).
20. Downward, J., Graves, J. D., Warne, P. H., Rayter, S. & Cantrell, D. A. Stimulation of p21^{ras} upon T-cell activation. *Nature* **346**, 719–723 (1990).
21. Favata, M. F. *et al.* Identification of a novel inhibitor of mitogen-activated protein kinase kinase. *J. Biol. Chem.* **273**, 18623–18632 (1998).
22. Mermoud, J. E., Cohen, P. & Lamond, A. I. Ser/Thr-specific protein phosphatases are required for both catalytic steps of pre-mRNA splicing. *Nucleic Acids Res.* **20**, 5263–5269 (1992).
23. Summerton, J. Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochim. Biophys. Acta* **1489**, 141–158 (1999).
24. Nasevicius, A. & Ekker, S. C. Effective targeted gene 'knockdown' in zebrafish. *Nature Genet.* **26**, 216–220 (2000).
25. Matter, N. *et al.* Heterogeneous ribonucleoprotein A1 is part of an exon-specific splice-silencing complex controlled by oncogenic signaling pathways. *J. Biol. Chem.* **275**, 35353–35360 (2000).
26. Dignam, J. D., Lebovitz, R. M. & Roeder, R. G. Accurate transcription initiation by RNA polymerase II in a soluble extract from isolated mammalian nuclei. *Nucleic Acids Res.* **11**, 1475–1489 (1983).

Acknowledgements We thank S. Stamm and O. Stoss for discussions and the gift of antibodies against SLM-2 and Sam68; S. Richard for the murine myc-Sam68 expression construct; S. Weg-Remers for luciferase plasmids; J. Sleeman and G. Dreyfuss for antibodies (2D5 and 4B10, respectively); U. Rahmsdorf and H. Olinger for technical assistance; and J. Valcárcel and I. Mattaj for advice on *in vitro* splicing and mRNA transport, respectively. This work was supported by the Deutsche Forschungsgemeinschaft.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to H.K. (e-mail: harald.koenig@itg.fzk.de).

Structure of the inositol 1,4,5-trisphosphate receptor binding core in complex with its ligand

Ivan Bosanac*, Jean-René Alattia*, Tapas K. Mal*, Jenny Chan*, Susanna Talarico*, Frances K. Tong*, Kit I. Tong*, Fumio Yoshikawa†, Teiichi Furuichi‡, Miwako Iwai‡, Takayuki Michikawa‡§, Katsuhiko Mikoshiba†§ & Mitsuhiro Ikura*

* Division of Molecular and Structural Biology, Ontario Cancer Institute and Department of Medical Biophysics, University of Toronto, 610 University Avenue, Toronto, Ontario, Canada M5G 2M9

† Laboratory for Developmental Neurobiology and Molecular Neurogenesis, Brain Science Institute, RIKEN, Saitama 351-0198, Japan

‡ Department of Molecular Neurobiology, Institute of Medical Science, University of Tokyo, Tokyo 108-8639, Japan

§ Calcium Oscillation Project, ICORP, Japan Science and Technology Corporation (JST), Tokyo, 108-0071, Japan

In a variety of cells, the Ca^{2+} signalling process is mediated by the endoplasmic-reticulum-membrane-associated Ca^{2+} release channel, inositol 1,4,5-trisphosphate ($InsP_3$) receptor ($InsP_3R$)¹. Being ubiquitous and present in organisms ranging from humans to *Caenorhabditis elegans*, $InsP_3R$ has a vital role in the control of cellular and physiological processes as diverse as cell division, cell proliferation, apoptosis, fertilization, development, behaviour, memory and learning². Mouse type I $InsP_3R$ ($InsP_3R1$), found in high abundance in cerebellar Purkinje cells,

is a polypeptide with three major functionally distinct regions: the amino-terminal $InsP_3$ -binding region, the central modulatory region and the carboxy-terminal channel region². Here we present a 2.2-Å crystal structure of the $InsP_3$ -binding core of mouse $InsP_3R1$ in complex with $InsP_3$. The asymmetric, boomerang-like structure consists of an N-terminal β -trefoil domain and a C-terminal α -helical domain containing an 'armadillo repeat'-like fold. The cleft formed by the two domains exposes a cluster of arginine and lysine residues that coordinate the three phosphoryl groups of $InsP_3$. Putative Ca^{2+} -binding sites are identified in two separate locations within the $InsP_3$ -binding core.

The recombinant construct encoding only residues 224–604 of mouse $InsP_3R1$ ($InsP_3R1_c$; Fig. 1) binds $InsP_3$ with high affinity ($K_d \approx 0.09$ nM)³. The crystal structure of mouse $InsP_3R1_c$ reveals that the protein forms an elongated L-shaped structure with two asymmetric domains oriented perpendicularly to each other (Fig. 2a). The N-terminal domain (residues 224–436) is composed of 12 β -strands and 2 single-turn helices, forming a globular barrel with approximate dimensions $35 \times 40 \times 30$ Å³ (Fig. 2a). In contrast, the C-terminal portion of mouse $InsP_3R1_c$ (residues 437–604) consists of eight α -helices, forming a bundle with dimensions $34 \times 26 \times 30$ Å³ (Fig. 2a). At the interface of these structurally distinct domains, hereafter referred to as the β -domain and the α -domain respectively, resides a deep cleft lined with basic residues. These residues are responsible for anchoring $InsP_3$ to the receptor (Fig. 2d), in a manner totally different from the phosphoinositides-binding modes observed in the pleckstrin homology (PH), ENTH (for 'epsin N-terminal homologue'), Tubby and FERM (for 'band 4.1, ezrin, radixin and moesin') domains^{4,5}. The conformation of

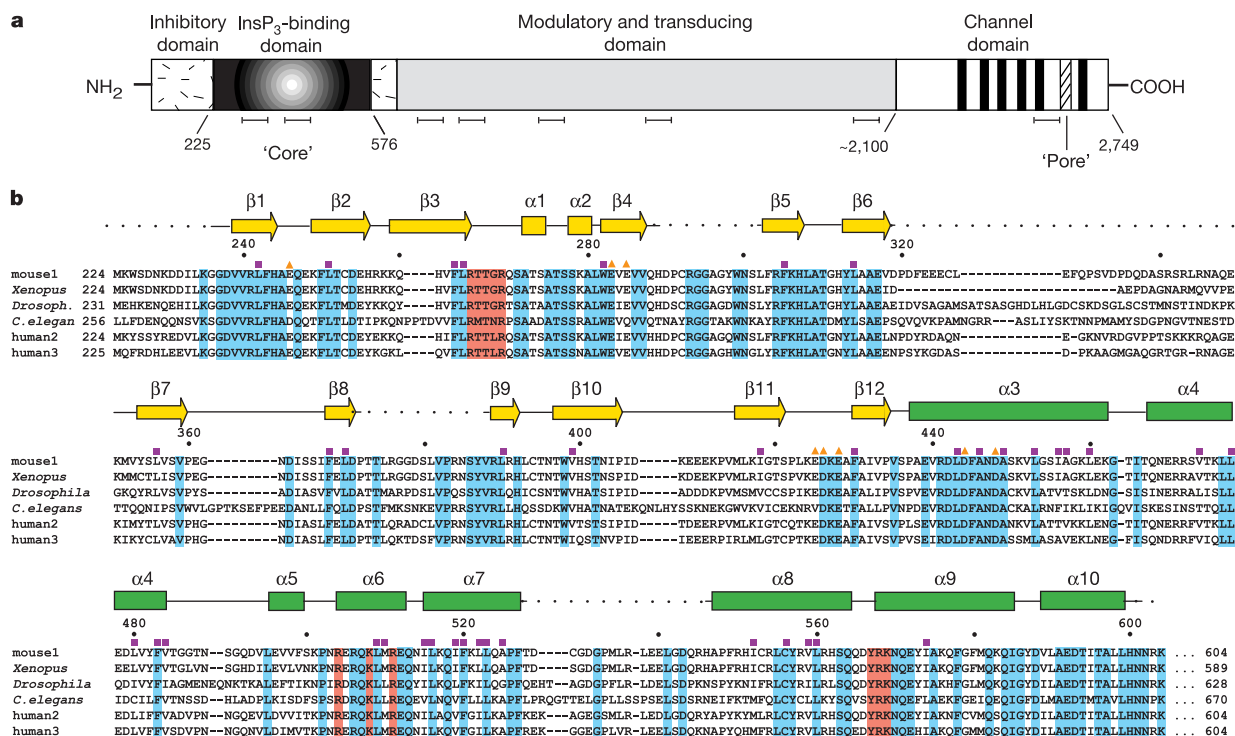


Figure 1 Overall domain architecture of $InsP_3R$ and sequence alignment of the $InsP_3$ -binding core between members of the $InsP_3R$ family. **a**, The four functional domains of $InsP_3R1$ with Ca^{2+} -binding sites (horizontal bars), putative membrane-spanning segments (vertical solid bars) and pore-forming region (diagonally striped box). **b**, Sequence alignment of the $InsP_3$ -binding core of $InsP_3Rs$: blue, conserved residues; red, $InsP_3$ -coordinating residues; purple squares, hydrophobic core residues; orange

triangles, Ca^{2+} -binding residues. β -Strands (arrows), α -helices (bars) and unidentified portions (dotted lines) are shown. The β -domain and the α -domain are coloured yellow and green, respectively. National Center for Biotechnology Information (NCBI) accession numbers for sequences: mouse1, ACMSIT; *Xenopus*, A40743; *Drosophila*, CAB51853; *C. elegans*, CAB45862; human2, XP_006747; human3, Q14573.

the receptor-bound InsP_3 molecule is similar to that found in the crystal structures of isolated InsP_3 and InsP_3 bound to a PH domain^{6,7}. However, the phosphorus–phosphorus distances differ between this study (8.4, 7.2 and 5.0 Å for P1/P4, P1/P5 and P4/P5 pairs) and others (8.1, 6.9 and 4.1 Å for isolated InsP_3 , and 8.2, 7.2 and 4.3 Å for InsP_3 bound to the PH domain)⁸, with P4 and P5 separation being significant.

The protein fold of the β -domain is known as the β -trefoil and is found in a number of diverse proteins, including interleukins-1 β and 1 α , fibroblast growth factors and mannose receptors⁹. Despite the low identity (14%) between mouse $\text{InsP}_3\text{R1}_c$ and the mannose-receptor β -trefoil, the structure of these two β -trefoil folds can be readily superimposed (root-mean-square deviation (r.m.s.d.) = 2.3 Å) (Fig. 3a). In the β -domain of mouse $\text{InsP}_3\text{R1}_c$, three of the six two-stranded hairpins (strand pairs β 1 and β 4, β 5 and β 8, β 9 and β 12) make up a barrel, and the other three (β 2 and β 3, β 6 and β 7, β 10 and β 11) are in a triangular array that caps the barrel, giving rise to a molecular arrangement with a pseudo-three-fold symmetry.

Structural comparison of the α -domain, with known folds using the DALI algorithm¹⁰, revealed a high homology to the armadillo repeat fold found in β -catenin¹¹, importins¹² and adenomatous polyposis coli¹³. The α -domain can be superimposed on the three consecutive armadillo repeats of β -catenin¹¹ (r.m.s.d. = 3.3 Å). Generally, an armadillo repeat consists of three α -helices with short interhelical linkers. In mouse $\text{InsP}_3\text{R1}_c$, helices α 3 and α 4 correspond to the second and third helices of the first armadillo repeat, and α 6, α 7 and α 8 constitute a perfect armadillo repeat 2 (Fig. 3b). The third repeat in mouse $\text{InsP}_3\text{R1}_c$ consists of α 9 and α 10 and lacks the third helix in the construct used for the present study.

As a result, the position of α 10 is disoriented relative to the canonical armadillo repeat. In the aforementioned proteins these armadillo repeats act as a protein–protein interaction motif¹³. In mouse $\text{InsP}_3\text{R1}_c$, the two relatively large, highly conserved surface areas are identified within the α -domain. One conserved surface (P-I) encompasses the amino-acid residues P502, F527, Y557, Q564, Q571, E572, A575, F578 and Q582 (Fig. 2d). The other conserved surface (P-II) contains a high proportion of charged amino acids: E283, V286, K306, Y313, R441, D444, F445, N447, D448 and R506 (Fig. 2e). These conserved regions rich in aromatic residues might serve as interaction sites for other parts of the receptor or unidentified cellular proteins. One obvious candidate for such a binding domain is the N-terminal InsP_3 binding suppressor domain (residues 1–225), which reduces the InsP_3 -binding affinity of the core domain from 90 pM to 45 nM (ref. 3).

Two different sequence-based analyses were used to search for the potential presence of additional armadillo repeats extending downstream of the α -domain. First, secondary structure prediction with PHD¹⁴ showed that the region spanning residues 460–1500 is predominantly α -helical. Second, mGenTHREADER¹⁵ analysis uncovered matches for possible armadillo repeats in regions within residues 760–1740 at ‘high’ confidence levels (E -values between 0.002 and 0.007). Armadillo repeats of this length (~1,200 residues including residues 440–604) would provide a long arm with a length of ~200 Å and a diameter of 35 Å, potentially connecting the InsP_3 -binding domain and the modulatory domain. This is consistent with a recent electron microscope analysis of native $\text{InsP}_3\text{R1}$, which showed a rod-like linker structure between the InsP_3 -binding region and the transmembrane channel core^{16,17}.

Highly basic amino acid residues spanning the two domains are

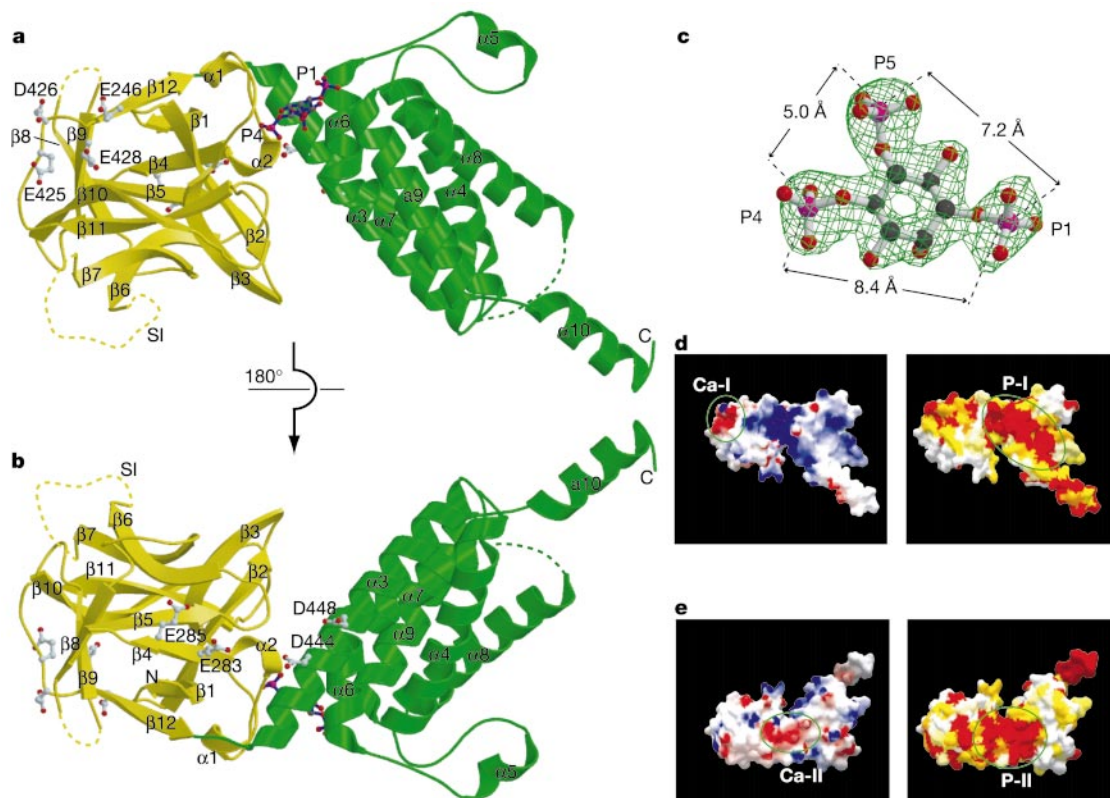


Figure 2 Structure of mouse $\text{InsP}_3\text{R1}_c$ in complex with InsP_3 . **a**, The β -domain (yellow) and the α -domain (green) with the InsP_3 molecule at the interface. Residues in the Ca-I and Ca-II sites and the splicing site (SI) are shown. **b**, View in **a** rotated by 180°. **c**, Experimental MAD electron density map at 1.7 σ around the InsP_3 molecule.

d, e, Molecular surface representations of mouse $\text{InsP}_3\text{R1}_c$ in the same orientations as **a** and **b**, respectively. Surface electrostatic potential (left panel with Ca-I and Ca-II sites) and conserved surface residues (right panel with P-I, P-II sites; identical residues are shown in red, and least-conserved residues in white).

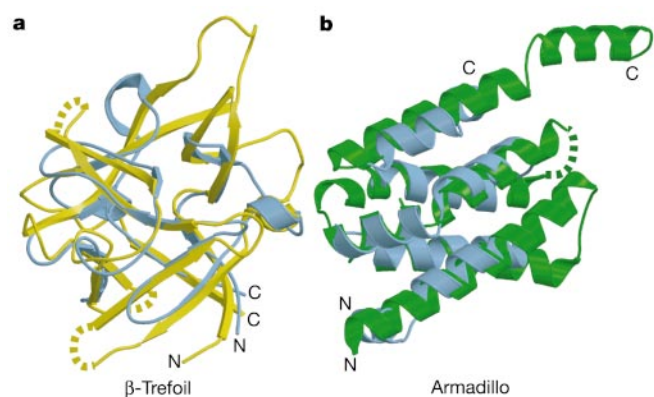


Figure 3 β -Trefoil and 'armadillo repeat'-like folds in mouse $\text{InsP}_3\text{R}_{1c}$. **a**, The β -domain (yellow) is superimposed on the cysteine-rich domain of the mannose receptor (blue) (accession number 1DQG, residues 4–128). The r.m.s.d. of the superposition is 2.3 Å. **b**, Superimposition of the α -domain (green) and β -catenin (blue) (accession number 3BCT, residues 269–368) with an r.m.s.d. of 3.3 Å. The size of one armadillo repeat of mouse $\text{InsP}_3\text{R}_{1c}$ is longer (a total of 62 amino acids in repeat 2) than those found in other proteins (typically 42 amino acids per repeat)¹³.

responsible for the extensive interactions of InsP_3 with InsP_3R . In addition to the interactions between protein side chains and phosphorus, six water molecules are directly involved in the formation of hydrogen bonds that bridge InsP_3 and the receptor (Fig. 4c). The coordination of the P1 and P5 groups predominantly involves residues from the α -domain (Fig. 4a), whereas the P4 group anchors to mouse $\text{InsP}_3\text{R}_{1c}$ by means of hydrogen-bond interactions mediated by β -domain residues (Fig. 4b). P4 is

involved in an extensive network of hydrogen bonds with residues R265, T266, T267, G268, R269 and K569. P5 also forms numerous hydrogen bonds with R265, R269, R504, K508, R511 and Y567. In contrast, P1 interacts only with R568 and K569 (Fig. 4). A less significant role for P1 has previously been suggested¹⁸. The hydroxyl groups of InsP_3 have a secondary role in binding to InsP_3R , which is consistent with earlier physico-chemical studies with monodeoxy, dideoxy and trideoxy derivatives of InsP_3 (ref. 19).

The current structure confirms the involvement of 9 out of 12 previously identified Arg/Lys residues as InsP_3 -coordinating residues, whose mutations to Gln drastically reduce or abolish InsP_3 -binding activity²⁰. Exceptions are the three amino acids R241, K249 and R506. Residing outside the InsP_3 -binding site, R241 in the β -domain forms a salt bridge with E439 in the α -domain, thereby stabilizing the inter-domain interaction. K249 and R506 are both located adjacent to the InsP_3 -binding site and form salt bridges with T266 and D442, respectively, thereby contributing to the structural integrity of both α - and β -domains. The structure further reveals the involvement of non-basic residues T266, T267, G268 and Y567 in InsP_3 coordination. As a result, we constructed four additional mouse $\text{InsP}_3\text{R}_{1c}$ mutants carrying T266A, T267A, Y567F or Y567A (Fig. 4d). The mutant T266A showed very little effect on InsP_3 binding, which was consistent with InsP_3 coordination via the backbone carbonyl group rather than the side chain (Fig. 4). The last three mutants all showed a drastic reduction in InsP_3 binding, confirming the importance of the side-chain functional groups of T267 and Y567 in coordination with InsP_3 .

Taylor and Marchant²¹ have demonstrated the importance of a direct coupling between InsP_3 and Ca^{2+} binding to the cytoplasmic portion of InsP_3R in the regulation of Ca^{2+} release channel activity. Previous truncation mutagenesis studies²² identified two Ca^{2+} -binding fragments of InsP_3R , one encoding residues 304–381 and

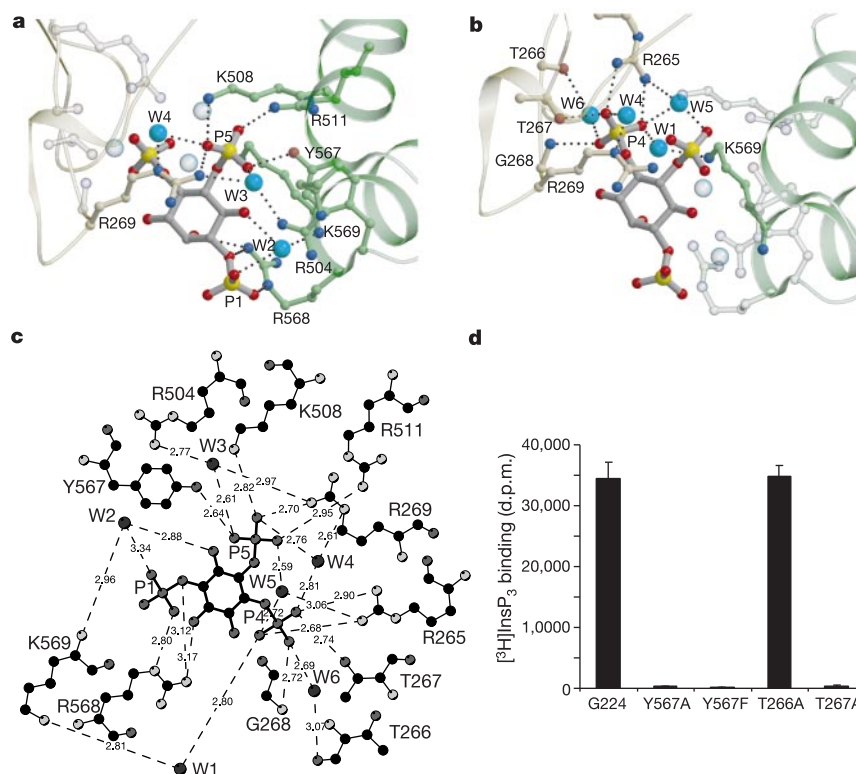


Figure 4 Coordination of InsP_3 by mouse $\text{InsP}_3\text{R}_{1c}$. Ribbon representation of the polypeptide chains of the α -domain (green) and the β -domain (yellow). Phosphates are shown in yellow, oxygens in red, nitrogens in blue, water molecules in cyan, and hydrogen bonds with dotted black lines. The coordination of P1 and P5 groups is shown in **a**, and

that of P4 in **b**, **c**. Two-dimensional schematic representation of the interaction of mouse $\text{InsP}_3\text{R}_{1c}$ with InsP_3 . **d**, Graph showing average specific InsP_3 -binding activity of site-directed mutant from a minimum of four independent experiments.

the other corresponding to residues 378–450. More recently, residues E425, D426, E428, D442 and D444 have been shown to be essential for Ca^{2+} coordination²³. These residues are part of the two surface acidic clusters identified in the present crystal structure of mouse $\text{InsP}_3\text{R}_{1c}$. The first site, Ca-I, is located in the β -domain and consists of residues E246, E425, D426 and E428 (Fig. 2d). The second site, Ca-II, located across the two domains, is composed of residues E283, E285, D444 and D448 (Fig. 2e). Interestingly, Ca^{2+} -binding site Ca-II overlaps with the conserved region P-II, suggesting that the binding of Ca^{2+} to this site is conformationally coupled with the aforementioned protein–protein interaction involving other protein domain(s). This finding, together with recent electron microscope studies of $\text{InsP}_3\text{R}^{16,17}$ and previous biochemical studies²¹, leads to a tempting speculation on the Ca^{2+} – InsP_3 coupling mechanism required for channel activation. The role of binding of InsP_3 to the core domain (residues 226–576) might include the release of a conformational constraint that prevents Ca^{2+} from binding to the receptor. The N-terminal InsP_3 binding suppressor region (residues 1–225) might be directly involved in this negative regulation of Ca^{2+} binding to the receptor, in addition to the modulation of InsP_3 binding affinity²⁰. It is equally possible that some other part of the InsP_3R or an unidentified cellular protein is involved in this InsP_3 – Ca^{2+} coupling mechanism. □

Methods

Overexpression and purification of InsP_3R_c

The mouse type 1 InsP_3R protein fragment encompassing the ligand-binding core, residues 224–604, was expressed as a glutathione S-transferase (GST) fusion protein (GST–mouse $\text{InsP}_3\text{R}_{1c}$) by polymerase chain reaction, cloning and sequencing as described previously^{20,24}. The protein was expressed in *Escherichia coli* strain BL21-CodonPlus-RIL (Stratagene) at 15 °C for ~20 h by induction with 0.5 mM isopropyl β -D-thiogalactopyranoside²⁴. GST–mouse $\text{InsP}_3\text{R}_{1c}$ was first purified with glutathione–Sephacrose (Pharmacia) followed by digestion overnight with thrombin to remove GST from the cutting buffer (20 mM Tris–HCl pH 8.4, 200 mM NaCl, 10% (v/v) glycerol and 20 mM dithiothreitol) at 4 °C with constant mixing. Cation-exchange chromatography (Fractogel EMD SO_3^- resin; EM Industries Inc.) and size-exclusion chromatography (Superdex 75; Pharmacia) were then performed to remove thrombin and any other impurities. Selenomethionine-labelled protein was produced by the technique involving the inhibition of methionine biosynthesis²⁵ and was confirmed by electrospray mass spectrometry. The purified protein was concentrated to 20 mg ml^{−1} in a crystallization buffer (15 mM Tris–HCl pH 8.0, 300 mM Na_2SO_4 , 2 mM tris-(2-carboxyethyl)phosphine and 50% molar excess InsP_3).

Crystallization and data collection

Initial crystals of mouse $\text{InsP}_3\text{R}_{1c}$ were produced at 295 K by the hanging-drop method by adding 2 μ l of protein complex to 2 μ l of well solution (100 mM MES pH 6.0, 20% polyethylene glycol (PEG) 3350, 0.2 M CH_3NO_2 and 2–5% dioxane). Crystals of the complex grew as extensive clusters of rods within 2 weeks. Series of microseedings were required to obtain single rod-type crystals with dimensions of $0.05 \times 0.1 \times 0.4$ mm³. Crystals were flash-cooled in crystallization buffer supplemented with 25% PEG 400. Multiwavelength anomalous dispersion (MAD) and native data were collected at 100 K on a BL44XU beam line at the SPring-8 Synchrotron facility and were processed with dT_{REK}²⁶. Crystals belonged to space group C22₁ with cell dimensions $a = 44.2$ Å, $b = 90.3$ Å and $c = 207.9$ Å, and contained one complex in the asymmetric unit. Radiation damage to the crystal prevented completion of the data set at λ_3 (remote wavelength) (Supplementary Table 1).

Structure determination and refinement

The initial positions of four out of six expected selenium atoms were determined by Solve²⁷ and refined with SHARP²⁸ to a figure of merit of 0.42. The position of the two missing selenomethionine residues (M224 and M535) were never identified because one is located at the N terminus of the protein fragment and the other in the $\alpha 7/\alpha 8$ loop region, which is not visible in the present model. Density modification and solvent flattening with the program CNS²⁹ increased the overall figure of merit to 0.93. The experimental map had continuous electron density for most of mouse $\text{InsP}_3\text{R}_{1c}$; the InsP_3 ligand was easily identifiable. Model building was performed with the program O³⁰ and refinement with CNS²⁹. The final model contained residues 236–288, 302–319, 351–372, 387–528 and 546–602, with 92.7% of residues in the most favoured regions of the Ramachandran plot and no residues in the disallowed region. The β -domain contains three crystallographically invisible loops, between strands $\beta 4/\beta 5$, $\beta 6/\beta 7$ and $\beta 8/\beta 9$ with the loop encompassing residues 320 to 350 ($\beta 6/\beta 7$) having a splicing site I. In total, 2,522 atoms with a mean temperature factor at 34.1 Å² were observed; 2,356 protein atoms ($B = 35.2$ Å²), 24 ligand atoms ($B = 28.6$ Å²) and 142 solvent atoms ($B = 36.2$ Å²). The R_{cryst} and R_{free} of the final model were 22.4 and 25.0, respectively.

InsP_3 binding assay

Site-directed mutagenesis of Y567A, Y567F, T266A and T267A was introduced into mouse $\text{InsP}_3\text{R}_{1c}$, and protein fragments were expressed in *E. coli*. Preparation of soluble protein from the cell pellets was performed as described elsewhere²⁴. Soluble protein (30 μ g) was diluted to 100 μ l with the binding buffer and incubated with 9.6 nM [³H] InsP_3 for 10 min at 4 °C in the absence and presence of 10 μ M InsP_3 to measure total and nonspecific binding, respectively. After the addition of 4 μ l of 50 mg ml^{−1} γ -globulin and 100 μ l of a solution containing 30% (w/v) PEG 6000, 1 mM EDTA and 50 mM Tris–HCl (pH 8.0 at 4 °C), the InsP_3 –protein mixture was incubated for a further 5 min at 4 °C. The protein–PEG complex was pelleted by centrifugation at 18,000g for 5 min at 4 °C. The resultant pellet was solubilized with 180 μ l of Solvable (Packard), then neutralized by the addition of 18 μ l of acetic acid. After mixing 5 ml of Atomlight (Packard), radioactivity of the samples was measured with a liquid-scintillation counter (Beckman). Specific binding activity was defined by subtracting nonspecific binding from total binding.

Received 15 August; accepted 29 October 2002; doi:10.1038/nature01268.

Published online 17 November 2002.

- Berridge, M. J., Lipp, P. & Bootman, M. D. The versatility and universality of calcium signalling. *Nature Rev. Mol. Cell Biol.* **1**, 11–21 (2000).
- Furuichi, T. & Mikoshiba, K. Inositol 1,4,5-trisphosphate receptor-mediated Ca^{2+} signaling in the brain. *J. Neurochem.* **64**, 953–960 (1995).
- Uchiyama, T., Yoshikawa, F., Hishida, A., Furuichi, T. & Mikoshiba, K. A novel recombinant hyperaffinity inositol 1,4,5-trisphosphate (IP_3) absorbent traps IP_3 , resulting in specific inhibition of IP_3 -mediated calcium signaling. *J. Biol. Chem.* **277**, 8106–8113 (2002).
- Overduin, M., Cheever, M. L. & Kutateladze, T. G. Signaling with phosphoinositides: better than binary. *Mol. Interventions* **1**, 151–159 (2001).
- Hamada, K., Shimizu, T., Matsui, T., Tsukita, S. & Hakoshima, T. Structural basis of the membrane-targeting and unmasking mechanisms of the radixin FERM domain. *EMBO J.* **19**, 4449–4462 (2000).
- Ferguson, K. M., Lemmon, M. A., Schlessinger, J. & Sigler, P. B. Structure of the high affinity complex of inositol trisphosphate with a phospholipase C pleckstrin homology domain. *Cell* **83**, 1037–1046 (1995).
- Hyvonen, M. et al. Structure of the binding site for inositol phosphates in a PH domain. *EMBO J.* **14**, 4676–4685 (1995).
- Hotoda, H. et al. Molecular recognition of adenophostin, a very potent Ca^{2+} inducer, at the D-myo-inositol 1,4,5-trisphosphate receptor. *Biochemistry* **38**, 9234–9241 (1999).
- Murzin, A. G., Lesk, A. M. & Chothia, C. beta-Trefoil fold. Patterns of structure and sequence in the Kunitz inhibitors interleukins-1 β and 1 α and fibroblast growth factors. *J. Mol. Biol.* **223**, 531–543 (1992).
- Holm, L. & Sander, C. Protein structure comparison by alignment of distance matrices. *J. Mol. Biol.* **233**, 123–138 (1993).
- Huber, A. H., Nelson, W. J. & Weis, W. I. Three-dimensional structure of the armadillo repeat region of β -catenin. *Cell* **90**, 871–882 (1997).
- Cingolani, G., Petosa, C., Weis, K. & Muller, C. W. Structure of importin- β bound to the IBB domain of importin- α . *Nature* **399**, 221–229 (1999).
- Peifer, M., Berg, S. & Reynolds, A. B. A repeating amino acid motif shared by proteins with diverse cellular roles. *Cell* **76**, 789–791 (1994).
- Rost, B., Sander, C. & Schneider, R. PHD—an automatic mail server for protein secondary structure prediction. *Comput. Appl. Biosci.* **10**, 53–60 (1994).
- Jones, D. T. GenTHREADER: an efficient and reliable protein fold recognition method for genomic sequences. *J. Mol. Biol.* **287**, 797–815 (1999).
- Hamada, K., Miyata, T., Mayanagi, K., Hirota, J. & Mikoshiba, K. Two-state conformational changes in inositol 1,4,5-trisphosphate receptor regulated by calcium. *J. Biol. Chem.* **277**, 21115–21118 (2002).
- Jiang, Q. X., Thrower, E. C., Chester, D. W., Ehrlich, B. E. & Sigworth, F. J. Three-dimensional structure of the type 1 inositol 1,4,5-trisphosphate receptor at 2.4 Å resolution. *EMBO J.* **21**, 3575–3581 (2002).
- Wilcox, R. A., Primrose, W. U., Nahorski, S. R. & Challiss, R. A. New developments in the molecular pharmacology of the myo-inositol 1,4,5-trisphosphate receptor. *Trends Pharmacol. Sci.* **19**, 467–475 (1998).
- Kozikowski, A. P., Ognyanov, V. I., Fauq, A. H., Nahorski, A. R. & Wilcox, R. A. Synthesis of 1D-3-deoxy-, 1D-2,3-dideoxy-, and 1D-2,3,6-trideoxy-myo-inositol 1,4,5-trisphosphate from quebrachitol, their binding affinities, and calcium release activity. *J. Am. Chem. Soc.* **115**, 4429–4434 (1993).
- Yoshikawa, F. et al. Mutational analysis of the ligand binding site of the inositol 1,4,5-trisphosphate receptor. *J. Biol. Chem.* **271**, 18277–18284 (1996).
- Marchant, J. S. & Taylor, C. W. Cooperative activation of IP_3 receptors by sequential binding of IP_3 and Ca^{2+} safeguards against spontaneous activity. *Curr. Biol.* **7**, 510–518 (1997).
- Sienaert, I. et al. Molecular and functional evidence for multiple Ca^{2+} -binding domains in the type 1 inositol 1,4,5-trisphosphate receptor. *J. Biol. Chem.* **272**, 25899–25906 (1997).
- Sienaert, I. et al. Localization and function of a calmodulin/apocalmodulin binding domain in the N-terminal part of the type 1 inositol 1,4,5-trisphosphate receptor. *Biochem. J.* **365**, 269–277 (2002).
- Yoshikawa, F. et al. High efficient expression of the functional ligand binding site of the inositol 1,4,5-trisphosphate receptor in *Escherichia coli*. *Biochem. Biophys. Res. Commun.* **257**, 792–797 (1999).
- Van Duyn, G. D., Standaert, R. F., Karplun, P. A., Schreiber, S. L. & Clardy, J. Atomic structures of the human immunophilin FKBP-12 complexes with FK506 and rapamycin. *J. Mol. Biol.* **229**, 105–124 (1993).
- Pflugrath, J. W. The finer things in X-ray diffraction data collection. *Acta Crystallogr. D* **55**, 1718–1725 (1999).
- Tervilliger, T. C., Kim, S. H. & Eisenberg, J. Generalized method of determining heavy-atom positions using the difference Patterson function. *Acta Crystallogr. A* **43**, 1–5 (1987).
- La Fortelle, E. D. & Bricogne, G. Maximum-likelihood heavy atom parameter refinement in the MIR and MAD methods. *Methods Enzymol.* **276**, 472–494 (1997).
- Brunger, A. T. et al. Crystallography and NMR system: A new software suite for macromolecular structure determination. *Acta Crystallogr. D* **54**, 905–921 (1998).
- Jones, T. A., Zou, J. Y., Cowan, S. W. & Kjeldgaard, M. Improved methods for binding protein models

in electron density maps and the location of errors in these models. *Acta Crystallogr. A* **47**, 110–119 (1991).

Supplementary Information accompanies the paper on *Nature's* website
(<http://www.nature.com/nature>).

Acknowledgements We thank A. Nakagawa, A. Miyazaki and K.T. Chong for their assistance at the beamline BL44XU at SPring-8, Japan; the staff at the X8-C and X25 beamline of Brookhaven National Laboratory and the BioCars beamline at Advanced Photon Source for their assistance; D. Jones and M. Swindells for mGenTHREADER analysis; and members of our division for discussions. This work was supported by grants from the Canadian Institutes of Health Research (CIHR) to I.B. and M.I., and by grants from RIKEN (to K.M.) and the Ministry of Education, Science, Sports, and Culture of Japan (to K.M. and T.M.). M.I. is a CIHR Investigator.

Competing interests statement The authors declare that they have no competing financial interests.

Correspondence and requests for materials should be addressed to M.I. (e-mail: mikura@uhnres.utoronto.ca). The atomic coordinates for InsP₃-bound mouse InsP₃R1_c have been deposited in the Protein Data Bank under accession code 1N4K. Macmillan Magazines Ltd., 2003 Nature Publishing Group London, UK 0028-0836/03/030280-00

erratum

The strength of Mg_{0.9}Fe_{0.1}SiO₃ perovskite at high pressure and temperature

Jiuhua Chen, Donald J. Weldner & Michael T. Vaughan

Nature **419**, 824–826 (2002).

On page 825 of this Letter, the equation should not have contained 'kern + 1' in the second term. The equation should read:

$$\text{FWHM}^2 = (2\varepsilon E)^3 + (Kh c / 2P \sin \theta_0)^2$$

□

Cellular solutions

The latest cell biology products, including a free CD-ROM.

CellTiter-Glo

Promega

www.promega.com

Shed more light on cell viability assays

Promega's CellTiter-Glo luminescent cell viability assay is a one-step, homogeneous protocol that combines sensitivity with ease and accuracy. The assay determines the number of viable cells in culture by quantifying the amount of ATP present — a marker of metabolically active cells. No cell washing, media removal or multiple pipetting steps are required, and measurements can be made as little as 10 minutes after addition of the reagent. As few as 15 cells can be detected. The assay is designed for 96- or 384-well formats and is suitable for automated, high-throughput screening of cell proliferation and cytotoxicity assays.

Smart Solutions for Cell-Based Screening

Beckman Coulter

www.beckmancoulter.com

CD-ROM for ArrayScan users

This CD-ROM contains searchable application notes, posters, and technical notes on the use of Beckman Coulter's ArrayScan high content screening system and HitKits reagent kits. It is aimed at researchers in academia and the pharmaceutical industry who are exploring the cell signal pathways using cell-based assays that analyse the response of live cells by generating multiple images and parameter data for each microplate well. The CD-ROM is available free of charge from www.beckmancoulter.com/hcsfreecd.

CellMap

Bio-Rad

www.cellscience.bio-rad.com

Novel confocal laser scanning microscopy systems for specific cell analyses

The cell science division of Bio-Rad is introducing two new application-directed products. CellMap ID and CellMap IC are novel confocal laser scanning microscopy systems. CellMap IC is designed for characterizing protein location in cells relative to nuclear and cellular structure. CellMap ID is designed for determining protein location in relation to the cell-cycle and for characterizing tissue morphology. CellMap IC and CellMap ID can be seen at Cell on 14–18 December in San Francisco.

Linear acrylamide

Amresco

www.amresco-inc.com

In line with protocols

Amresco's linear acrylamide, 5 mg/ml, is a co-precipitant that helps improve the yield of nucleic acids in the precipitation steps of purification protocols. The polyacrylamide solution is a neutral carrier and is not derived from a biological source. As a result, it is free from the potential of containing nucleic acid or nuclease contamination, and does not interfere with A_{260}/A_{280} readings. The solution remains stable when stored cold (4 °C) and is supplied in a five-pack of 1-ml vials designed for ease of dispensing.

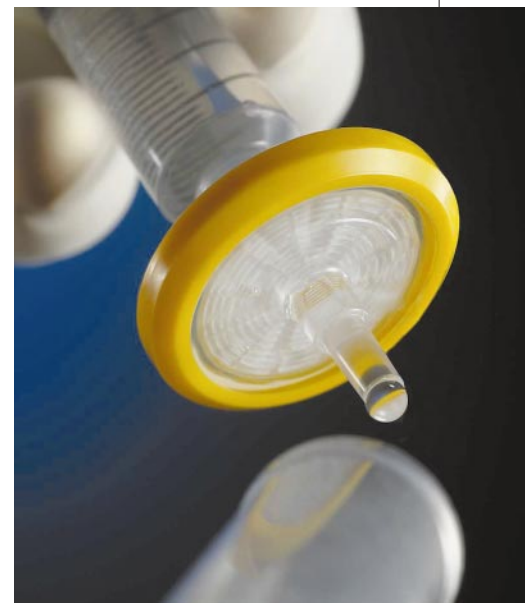
Antibody for *Xenopus*

Serotec

www.serotec.co.uk

Foreign correspondents

This monoclonal antibody from Serotec specifically recognizes a 65 kD *Xenopus* protein, shown to be homologous to human ribophorin I. All newly synthesized proteins that are destined for the cell surface and secretion are folded and assembled in the rough endoplasmic reticulum (RER). In mammals, researchers have identified an oligomeric complex — oligosaccharyl-transferase (OST) — that has enzyme activity and affects the glycosylation of newly synthesized polypeptides in the RER. The OST complex is reported to be composed of four ER-specific membrane proteins, including ribophorin I, ribophorin II, OST48 and Dad1. The mechanisms by which ribophorin I is located to the RER and the role of the protein in translation and post-translational modifications have attracted interest in recent years.



Millipore offers faster filtration.

Millex 33 mm syringe filter

Millipore

www.millipore.com

More on the surface

These filters have a surface area 20% larger than that of filters with a 25 mm diameter, thus increasing flow rate and throughput. They also have a maximum housing pressure of 150 psig (10 bar), allowing users to accelerate the overall filtration process. The units have a polyvinylidene fluoride membrane for low protein binding, and are available with pore sizes of 0.1, 0.22 and 0.45 µm.

DUALhybrid

Dualsystems Biotech

www.dualsystems.com

Customized yeast two-hybrid screening service

Dualsystems Biotech provides in-vivo identification of protein–protein interactions. The company performs customized and optimized screenings based on the yeast two-hybrid system and variations thereof. Results are obtained within three months.

PicoPure protein extraction buffer

Arcturus

www.arcturus.com

Pure cell proteomics

PicoPure is a general-purpose chaotropic buffer validated for protein extraction from laser capture microdissected (LCM) cells. Complete protocols for protein extraction from samples on CapSure LCM caps are included, along with 10 ml of buffer.



Order your free CD from Beckman Coulter.

Finnpipette Focus

Thermo Life Sciences www.thermo.com

For delicate manoeuvres

This ergonomically designed pipette features a choice of small, medium and large snap-on handles, a short tip to enhance precision, fast volume setting, advanced fine volume adjustment and superior blow-out. The single-channel variable volume pipettes are available in a selection of 10 volume ranges, from 0.3 µl to 10 ml.



Focus on ergonomics: Finn timer is designed to offer comfort and control.

TranSignal array

Panomics www.panomics.com

TF–TF interactions assessed in a single experiment

This array allows assessment of interactions between multiple transcription factors (TFs) in a single hybridization experiment. Instead of directly measuring interactions between proteins, it employs a sensitive, indirect approach based on the fact that every TF has its own responsive DNA binding element, or cis-element. TF–TF interactions can be assessed by looking at the bound cis-elements — the interactions can be indirectly visualized when corresponding cis-elements are hybridized to the membrane. The array is spotted with 96 TFs.

ArrayScan 4.0

Cellomics www.cellomics.com

High content screening reader

The ArrayScan reader allows users to extract vital data on the relative distribution and activity of cellular constituents combined with morphometric measurements in fixed samples. Version 4.0 incorporates an automated optical bench, expanded filter combinations, additional objective options and optimized computer hardware and software. A 'protocol assignment' feature allows multiple users to carry out different high content screening assays in the same unattended run. Expanded six-colour capability enables wider reagent selection, and Cellomics promises a 40% reduction in plate scan times.

Roll & Grow

Qbiogene www.qbiogene.com

Plating beads

Roll & Grow plating beads offer an alternative to the traditional 'bent glass rod' method of preparing agar plates for incubation. Cells are pipetted onto the plate, and the plating beads are then added and rolled around the plate using a back-and-forth tilting motion. The movement of the beads spreads the yeast or bacterial cells evenly across the entire surface of the plate. Once spreading is complete, the beads are discarded and the plate can be incubated.

Filters for flow cytometry

Omega Optical www.omegafilters.com

To dye for

This line of filters has been designed to accommodate the growing number of lasers and fluorescent detection channel instruments. While three- or four-fluorochrome analysis has become standard, simultaneous detection of 10 colours is now possible. These filters maximize multiple dye discrimination. They are manufactured to fit most makes of research and clinical instruments.

Fireboy adapter

Integra Biosciences
www.integra-biosciences.com

Step on the gas

Designed for use with the company's Fireboy Plus and Fireboy Eco mobile Bunsen burners, this adapter/safety stand accommodates common larger gas cartridges, including Camping Gaz CP 250 and other butane (250) cylinders. It ensures that larger gas cartridges remain in the correct position and enhances safety. Applications include sterilization and sample heating.

Expression system

Invitrogen www.invitrogen.com

Undivided attention

ViraPower allows reproducible and stable expression of a gene of interest without the need for a round of replication, even in

non-dividing mammalian cells. Reproducible analyses of long-term, high-level expression in the entire population of any mammalian cell type are possible, including cell types that challenge standard transfection or other viral transduction experiments.

Pro ID Software

Applied Biosystems
www.appliedbiosystems.com

Automated protein identification in proteomics

Pro ID software has been developed for automated protein identification in complex biological samples when used for analysis of tandem MS data from proteomics experiments. Pro ID software runs on the API QSTAR Pulsar Hybrid LC/MS/MS system. It incorporates a proprietary algorithm for rapid protein identification. The novel Interrogator database search engine provides high-throughput protein identification from uninterpreted tandem MS data. Besides finding specific (expected) modifications, such as oxidation or phosphorylation, Pro ID software also performs error and modification-tolerant searches, using a novel 'zone modification' feature.

These notes are compiled in the Nature office from information provided by the manufacturers. Press releases to be considered for publication can be sent to newproducts@nature.com.

Overview of the Alliance for Cellular Signaling

Participating investigators and scientists of the Alliance for Cellular Signaling*

*A full list of AfCS participating investigators and scientists appears at the end of this paper

The Alliance for Cellular Signaling is a large-scale collaboration designed to answer global questions about signalling networks. Pathways will be studied intensively in two cells — B lymphocytes (the cells of the immune system) and cardiac myocytes — to facilitate quantitative modelling. One goal is to catalyse complementary research in individual laboratories; to facilitate this, all alliance data are freely available for use by the entire research community.

The availability of complete genomic sequences and the emerging technologies of proteomics inspire confidence that complex biological phenomena and systems can be understood completely. Such optimism is strengthened by rapidly expanding capabilities to undertake global analyses of biological systems, including the capacity to manipulate gene content and expression in mammalian cells and organisms, to detect protein–protein interactions, and to quantify the activities of macromolecules *in vivo*. Such understanding of complete systems implies (and demands) the capacity to predict quantitatively the altered behaviour of these systems that results from their genetic or environmental (including pharmacological) perturbation. We can foresee, rather than simply imagine, the construction of a virtual cell, and this accomplishment will eventually have an enormous impact on biomedical and pharmaceutical science.

We have launched a major new research initiative called the Alliance for Cellular Signaling (AfCS; www.signaling-gateway.org): a multi-disciplinary, multi-institutional consortium to study cellular signalling systems. The overall goal of the AfCS is to understand as completely as possible the relationships between sets of inputs and outputs that vary both temporally and spatially. The same goal, stated from a slightly different perspective, is to understand fully how cells interpret signals in a context-dependent manner. How does a cell respond appropriately to one voice when it must listen simultaneously to many, and how does it alter this response in the context of other concurrent or recent signalling events? Answers to these questions will require identification of all the molecules that comprise the various signalling systems, assessment of time-dependent information flow through the systems in both normal and pathological states, and reduction of the mass of detailed data into a set of interacting models that describe cellular signalling.

The organization of the alliance was catalysed by an announcement by the National Institute of General Medical Sciences (NIGMS) of plans to fund large-scale collaborative projects to “enable the solution of major problems in biomedical research and to facilitate the next evolutionary stage of integrative biomedical science.” NIGMS calls these ‘glue’ grants. The strategy of the AfCS includes a ten-year research plan with an annual budget of about US\$10,000,000. We have obtained the necessary funding from NIGMS, the National Institute of Allergy and Infectious Diseases, the National Cancer Institute, a consortium of pharmaceutical companies (Eli Lilly, Aventis, Johnson and Johnson, Novartis and (briefly) Chiron) and private donors (The Merck Genome Research Institute, the Agouron Institute and an anonymous philanthropist).

We plan to use these funds to answer overarching questions that cannot be approached by individual laboratories (Table 1). We are keenly aware of the fact that by acting alone, the alliance can acquire only a small fraction of the necessary information. A critical part of our strategy is thus to catalyse complementary research throughout the signalling community. We provide leads for pursuit by individual

laboratories by disseminating all of our data publicly and promptly via the Internet. Information is released as soon as it has been verified, and neither AfCS investigators nor our sponsors have access to data before it is available to the public. We do experiments, but, importantly, we are ourselves an experiment in how collaborative science might be accomplished.

Organization of the AfCS

The so-called participating investigators of the AfCS comprise a group of around 50 scientists at about 20 different institutions, predominantly in the United States, with a few representatives from Canada and the United Kingdom. Participation is enhanced by communication and data sharing by videoconference using Internet2. (Internet2 is a university-led consortium whose goal is to create a leading-edge network capability for the national research community.)

The participating investigators are members of various committees or are directors of AfCS laboratories (Fig. 1; and see list of investigators at the end of this paper). A critical interaction is that between the laboratory directors and two ‘system committees’, representing the B lymphocyte and the cardiac myocyte (the two cells of primary interest to the AfCS). The system committees are charged with experimental planning and data interpretation, while the laboratory directors design and implement the experimental plan.

There is complete separation between the AfCS laboratories and the individual laboratories of the participating investigators. Participating investigators do not receive funding for their own laboratories, with the exception of modest funds that are provided for a few so-called ‘bridging projects’ (aimed at technology development). The eight AfCS laboratories are staffed by dedicated PhD scientists and technicians. The relatively small number of AfCS laboratories in which work is done minimizes experimental sources of variance as data are collected systematically.

Membership in the AfCS constitutes another level of participation in the venture. The role of the more than 1,500 members is described below.

Selecting the cells to be studied

We are committed to the tenet that much data must be gathered under standardized conditions on a small number of systems to permit the desired depth of quantitative understanding of signalling system behaviour. We have limited ourselves to two cell types — the B lymphocyte and the cardiac myocyte. The two cell types were not easy choices, indeed both pose enormous challenges, and agreement was reached only after a great deal of discussion. The primary criteria for choice of cells were that they should be mammalian with a sequenced genome, a criterion that drove us to the mouse; they should be nonmalignant and euploid, which led us to select cells that were as normal as possible and whose signalling systems had not been co-opted by malignant transformations; and they should be of requisite mass and homogeneity, offer

Table 1 Central questions of the AfCS

Question	Experiment	New technologies required
1. How complex is signal processing in cells? Considering the set of ligands for cell-surface receptors as a potential combinatorial code of inputs, how much of this input complexity can a cell uniquely decode at its outputs? In other words, how big is the AfCS problem, and can we experimentally bound the complexity of cellular signalling?	Systematic single- and double-ligand screens to classify output responses and determine degree of functional crosstalk between transduction pathways.	Analytic methods to classify and compare multi-dimensional data for different ligand combinations.
2. What is the structure of the whole signalling network? Is the connectivity sparse or dense?	Wholesale mapping of physiologically relevant protein–protein and small-molecule–protein interactions.	High-throughput assays for inter-molecular interaction <i>in vivo</i> , especially in response to ligand stimulation.
3. How much does network topology constrain signal processing capability? How much of function is specified by the nature of connections rather than by the specific biochemical constants of individual activities? Can a functional module be defined and recognized simply by network topology?	Perturbation methods (gain-of-function and loss-of-function) coupled with high-throughput functional assays to test the role of potential network modules.	Methods for reliably perturbing the concentration (or specific activity) of signalling proteins <i>in vivo</i> , both singly and in combinations.
4. Can functional modules be abstracted mathematically? Can we make physical models of the network modules that explain and predict the input–output relations of signalling subsystems? Can we identify the basic computations being done by these modules? What is the ‘device physics’ of biological signalling components?	Mathematical modelling of network modules coupled with experimental testing.	New mathematical methods to model the behaviour of signalling systems. Approaches to improve the interface between theorists and experimentalists.
5. What are the dynamics of the signalling network? Can we visualize how perturbations to the signalling network (such as ligand application) propagate through the network as changes in functional activities? This represents the flow and transformation of information during signalling events.	Follow the kinetics of signalling through a set of nodes in the signalling network that comprise a representative sampling of the sites of information processing.	Methods for monitoring protein activities with high temporal and spatial resolution in cells.
6. Why is the network the way it is? What general thermodynamic principle operates on evolved systems to make the observed solutions to information processing the chosen ones? What is being optimized so that the signalling network looks and behaves the way it does?	Unknown for now.	Unknown for now.

the potential to manipulate gene expression, and possess interesting biological properties regulated by interacting signalling systems.

The significant advantages and the disadvantages of these cells are described in the accompanying articles (pages 708–710 and 712–714). There is no single cell type that satisfies all of our desires — it is difficult to find all wanted attributes and bring all necessary technologies to a single system — and it remains to be seen if we have struck a set of manageable compromises.

Experimental strategy

Our goal is to understand the cellular signalling networks that convert a set of time-varying inputs into physiologically relevant outputs. This requires identification and localization of all the relevant molecules, definition of the physical interactions between these molecules, and quantitative assessment of the physiological significance of these interactions as information flows through the system. Furthermore, it is critical that this flow of information be evaluated under conditions where interactions between combinations of inputs can be assessed, as this is the only way to probe the level of complexity of the system. These conceptual requirements dictate three broad and temporally ordered (but overlapping) experimental approaches, outlined in Table 2.

The parts list

What signalling proteins are expressed in the cells of interest, where are they located within these cells, and do they move within these cells in response to the flow of regulatory information? To answer these questions we must conduct broad screens to detect interactions between proteins and to identify players that do not yet appear on the AfCS protein list. We are conducting yeast two-hybrid screens using AfCS proteins as bait in collaboration with Myriad Genetics. Results are posted on the alliance website as they are received and should represent a valuable source of leads for pursuit by individuals in the signalling research community. We hope to launch other protein- and DNA-based screens for relevant interactions in the near future.

Ligand screens

We need to know what ligands will activate signalling pathways in the chosen cells, what combinations of ligands will reveal interactions between signalling pathways, and what downstream responses are reliable indicators of signalling activity. The single and multiple ligand screens are designed to provide this information. B cells and cardiac myocytes are being exposed to a large number of ligands chosen according to information from the literature. We are sampling cellular responses to each ligand with a collection of assays that should reveal the

nature and breadth of the response, including second messenger concentrations, alterations in protein phosphorylation, patterns of gene expression and, in time, concentrations of many cellular lipids. We anticipate compilation of a list of 20 or more active ligands for each cell type. The single ligand screen will tell us if any one ligand is truly the same as another (for example, how much do the agonists that work on the distinct G_i -coupled chemokine receptors of B cells differ from each other) and will reveal responses to be targeted for careful quantification when we return to measure information flow through the entire system.

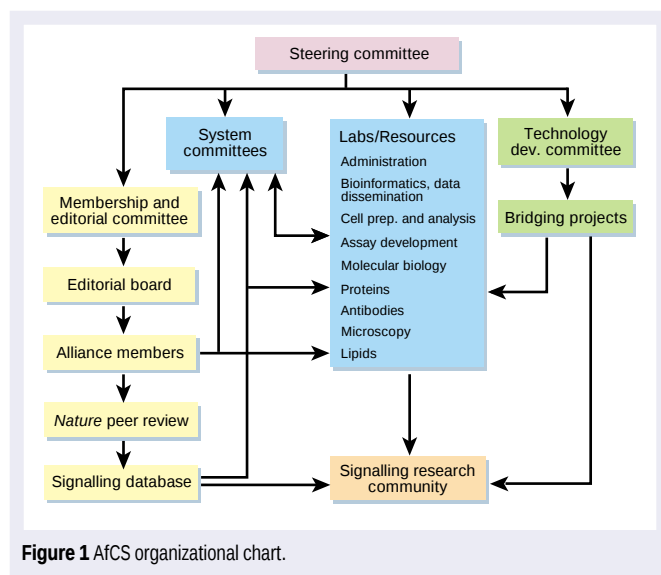
Measurements of responses to combinations of ligands will reveal interactions between pathways. Interaction is revealed by energetic (logarithmic) non-additivity. We will thus evaluate the additivity of the log fraction (or multiple) of the output. The combinations of inputs that yield the most extensive deviations from energetic additivity will be carried forward for rigorous analysis. We cannot fully sample the multidimensional space of two-way ligand combinations (various concentrations, time points, order of additions, and so on), to say nothing of higher-order combinations. Additionally, we have no idea of the density of interactions that we might uncover, but some estimate of this parameter will provide the first at least semi-systematic estimate of the level of complexity of cellular signalling systems.

Perturbational analysis

With parts lists and ligands in hand, we will undertake a perturbational strategy to construct an epistasis map. Such approaches are as old as pharmacology and genetics, where they were born. Most investigators would likely agree that pharmacological manipulation of a system (be it called chemical genetics or not) has more to offer (in terms of speed and reversibility) when cell-permeant chemicals of high specificity are available. But the number of adequate chemical perturbants is unfortunately small, and we must rely heavily on nucleic acid-based techniques to manipulate gene expression to achieve either loss or gain of function. RNA interference is particularly promising, and successful application of this technology would greatly improve the quality of our effort. Antisense approaches are also being pursued. Just as we will apply combinations of ligands to our experimental cells, it will become equally important to use combinations of perturbants.

Quantitative determination of information flow

The final experimental phase of the analysis will be the quantitative determination of functional interactions among the signalling molecules. This ambitious undertaking will start when the system committees decide that the identification of new signalling



molecules and their interactions is approaching saturation. The data from this effort will allow us to attain our goal of performing the first molecular analysis of how a signalling network actually works.

For this goal, it will almost certainly be inappropriate to adopt the traditional reductionist approach of simply measuring *in vitro* all the K_d , K_m and V_{max} values, cooperativity constants, rate constants and molar concentrations for all molecules involved. One reason is that it is unlikely that any artificial solution (medium) will allow approximation of the physiological magnitude of these values. Multiple specific interactions of a single molecule in the cell will also be impossible to duplicate *in vitro*. And determination of the vast array of values that the strict reductionist approach demands will probably be technologically impossible in the foreseeable future, and their manipulation to yield a model of cell function would be daunting.

An alternative approach to quantifying the matrix of regulatory interactions in the cell is to measure functional interactions on essentially arbitrary scales for each molecule that we test, and then to link fractional output (fractional activation, activity and binding) of one molecule with its fractional effect on those with which it interacts. In general, we will stimulate and/or inhibit the activity of each molecule p over a defined range and measure consequent changes in the activities (or other parameters) of all molecules q_n with which p interacts directly in a functionally important way. By asking how much of the maximal activity of p does it take to produce fractional activation of $q_1, q_2, \dots, q_n$, we hope to arrive at a set of linkage functions that should in principle be determined by fractional activation of receptors by ligands and be interpretable as fractional physiological outputs.

What are the experimental parameters that will give us these fractional activities? Each will depend on the type of molecule measured. For Ca^{2+} , concentration is adequate, and appropriate assays are

already available. For a protein phosphatase, activity will vary with substrates, and fractional activation towards one substrate will not necessarily equal that towards another, as modulation can change specificity. Conceptually, any 'reporter' will do, but reporters must be closely linked to the molecules whose activities they describe or their information will be degraded (filtered) by other interactions in the signalling network. Development of appropriate *in situ* signalling assays is one of the more rapidly evolving technologies, and the alliance will draw on the leaders in this field. Spectroscopic outputs are clearly preferable, and product traps for enzymes that synthesize stable, low-molecular-weight molecules are also desirable.

Data modelling and network analysis

The alliance is built on two independent developments that have become prominent during the past decade: the success of molecular biology in developing tools to identify the component parts of biological systems and their interactions, and the development of mathematical and computational tools for the analysis and design of complex engineering systems.

The first of these advances provides measurement technology to yield quantitative assessments of the state of components of biological networks (for example, transcripts, small ions, labelled proteins and cell shapes) that begin to equal the tools used for the analysis of other engineering networks. In fact, the early stages of the alliance's efforts are devoted to the further identification of the components of cellular signalling networks and the quantitative measurement of their input–output behaviours and regulation. But appreciation of the biological implications of these measurements will be largely impenetrable without data analysis and modelling tools designed specifically for biological network systems. Thus the final stages of the alliance's activity must address the 'complex systems' aspects of biology, building on methods from systems engineering, computer science, control theory, circuit design and dynamical systems.

The current charge of the alliance's network analysis team is to produce the necessary theoretical and computational framework to achieve these final goals. This requires close interactions with the database and informatics teams to create data structures and tables appropriate for network analysis, and with the system committees to assure that the experimental procedures yield data that are appropriate for network prediction and analysis. We must create tools for organization and analysis of these data and then build static and dynamic models of system function that can be validated experimentally. These models can then be used to design practical biomedical strategies for control of these pathways and their associated disease states. Thus, this team must forge the necessary tight cycle of experiment, theory and computation that is the hallmark of the analysis and engineering of networks in other fields of inquiry, and the alliance has been structured to permit this type of interaction.

Establishment of a signalling database

The AfCS will sponsor and coordinate the establishment of a signalling database that will be an invaluable resource for the entire signalling

Table 2 An overview of the experimental strategy of the AfCS

Identify and localize the relevant molecules	Define and expand the molecule list	Define the flow of information
1. Define basic scope, including pathways of interest and an initial list of molecules. (I)* 2. What molecules are present? Large-scale transcriptional profiling. (I) 3. Quantify proteins of interest (using immunoblotting, ligand binding). (I) 4. Localize proteins of interest (by microscopy, subcellular fractionation). (I) 5. Monitor movements of proteins, preferably quantitatively and in real time (for example, using GFP fusions); also a quantitative intermediate endpoint. (I)	1. Broad screens to detect interactions between proteins. The goal is to define and expand the list of molecular players, and to identify 'all' of the interactions among the components. The research community at large will define the physical and physiological validity of these interactions. The alliance will define the further physiological validity and the quantitative physiological significance of these interactions in its two cellular systems. (I)	1. Choose the inputs using broad assays to eliminate redundancy and detect interactions. (I) 2. Define intermediate endpoints. (I,L) 3. Choose the outputs. (I,L) 4. Measure intermediate endpoints and outputs as a function of inputs (for example, time, concentration, combinations) while altering functional concentrations of components (the molecule list, singly and in combinations) pharmacologically and 'genetically'. (L) 5. Define quantitative relationships between activity states of interacting components; assays in real time <i>in vivo</i> and snap shots of intracellular activity. (L) 6. Develop and test models. (L)

*I, work to be done initially; L, work that must be done later.

research community. The central element of this database will be the 'Molecule Pages', which will aim to represent every protein on the AfCS list (see article on pages 716–717). These will be standardized documents that each contain a wealth of literature-derived information about a given molecule; additional information from the alliance laboratories will be incorporated in time. Linkages will be provided between Molecule Pages to maps of signalling pathways as they are developed and refined, to the original literature, and to other databases.

Molecule Pages will be prepared (and updated) by AfCS members and will be peer reviewed, the entire process being supervised by an editorial board and *Nature*. We believe strongly that the collection and interpretation of the considerable amount of information about a given protein and its entry into a complex object-relational database is a substantial scholarly effort — at least as important as a conventional written review because of the powerful nature of the database itself. Given peer review of these Molecule Pages and the important collaboration between the AfCS and Nature Publishing Group, we believe that AfCS Molecule Pages should and will have the cachet of a major publication in *Nature* or a similar high-profile journal. Members will be chosen particularly for their expertise about specific molecules and pathways, and applications for membership have been and will continue to be encouraged publicly via our website.

Socialistic aspects of the AfCS concept

The alliance is itself an experiment in collaborative science. There have been relatively few large-scale collaborations in basic biology. It is a young science that has flourished to date in the hands of individual investigators, directing relatively small laboratories. But the scientific questions posed by the AfCS are not suited to the traditional distributed approach to experimental biology.

Are there impediments to large-scale collaboration? There are significant nuisances, but they pale by comparison with the potential gain. Obvious difficulties imposed by distance can be largely circumvented with modern communication systems; the AfCS does not function without the Internet. Human and career issues are significant — scientists are no less egocentric or competitive than others. We may be endowed with exuberant senses of curiosity, but we can lay no collective claim to sainthood. In addition, the academic reward system compounds the problem, and the AfCS does not provide a wealth of conventional academic rewards.

It is essential that we distribute data publicly and promptly via the Internet; publications of experimental observations (but not analyses) are thus largely superfluous. The vast majority of AfCS funding is spent in the AfCS laboratories to optimize precision and standardization of data collection. The classical promotion and tenure system needs adjusting to recognize an individual's contributions to group efforts if large-scale biology is to come of age. Physicists have learned the art of large-scale collaboration because of collective needs for extremely expensive facilities. Biologists will learn the same art because of the overwhelming complexity of the problems to be solved.

Our challenge is substantial, and it can be met only by the coordinated and cooperative efforts of many, in roles ranging from preparation of Molecule Pages by our members, to donation of significant amounts of time and energy by our participating investigators, to full-time employment by AfCS laboratory scientists and technicians. These contributions will benefit the entire field enormously and expand the opportunities for everyone. Those involved with this project see the need clearly and bring their own motivation to the project. We are excited by the idea that key questions about cellular information processing can be answered by a collective approach and that such understanding will produce new paradigms to guide

further studies in individual laboratories. The alliance offers this opportunity in a unique way, but only to those who are willing to donate a part of themselves to the group. We need socialistic science practised by strong individualists. Practical? We believe it is. □

doi:10.1038/nature01304

Acknowledgements

This article was written by R. Taussig, R. Ranganathan, E. M. Ross and A. G. Gilman (University of Texas Southwestern Medical Center) on behalf of the participating investigators and scientists of the AfCS. R.R. is an Associate Investigator of the Howard Hughes Medical Institute.

Participating investigators and scientists of the AfCS

Steering Committee: Alfred G. Gilman¹ (Chair), Melvin I. Simon⁶ (Vice Chair), Henry R. Bourne⁹, Bruce A. Harris¹³, Rochelle Long¹⁴, Elliott M. Ross¹, James T. Stull⁷, Ronald Taussig⁷

B Lymphocyte System Committee: Henry R. Bourne⁹ (Chair), Adam P. Arkin¹⁵, Melanie H. Cobb¹, Jason G. Cyster¹⁰, Peter N. Devreotes¹⁶, James E. Ferrell¹⁷, David Fruman²⁹, Michael Gold²¹, Arthur Weiss¹¹

Cardiac Myocyte System Committee: James T. Stull⁷ (Chair), Michael J. Berridge²², Lewis C. Cantley²³, William A. Catterall²⁴, Shaun R. Coughlin¹², Eric N. Olson⁵, Temple F. Smith²⁵

External Advisory Committee: Joan S. Brugge²⁶ (Chair), David Botstein¹⁸, Jack E. Dixon²⁷, Tony Hunter²⁸, Robert J. Lefkowitz²⁹, Anthony J. Pawson³¹, Paul W. Sternberg⁷, Harold Varmus³²

Bioinformatics and Data Coordination Laboratory: Shankar Subramaniam³³ (Director), Robert S. Sinkovits³³, Joshua Li³³, Dennis Mock³³, Yuhong Ning³³, Brian Saunders³³

Cell Preparation and Analysis Laboratory: Paul C. Sternweis¹ (Director), Donald Hilgemann² and Richard H. Scheuermann³ (Assoc. Directors), Dianne DeCamp¹, Robert Hsueh¹, Keng-Mean Lin¹, Yan Ni¹

Laboratory for Development of Signaling Assays: William E. Seaman^{10,11,34} (Director), Paul C. Simpson^{11,12,35} (Co-Director), Timothy D. O'Connell³⁵, Tamara Roach³⁴

Molecular Biology Laboratory: Melvin I. Simon⁶ (Director), Sangdun Choi⁶, Pamela Eversole-Cire⁶, Iain Fraser⁶

Protein Chemistry Laboratory: Marc C. Mumby¹ (Director), Yingming Zhao⁴ (Consultant), Deirdre Brekken¹, Hongjun Shu¹

Microscopy Laboratory: Tobias Meyer¹⁹ (Director), Grisca Chandy¹⁹, Won Do Heo¹⁹, Jen Liou¹⁹, Nancy O'Rourke¹⁹, Mary Verghese¹⁹

Antibody Laboratory: Susanne M. Mumby¹ (Director), Heping Han¹

Lipidomics Laboratory: H. Alex Brown³⁶ (Director), Jeffrey S. Forrester³⁶, Pavlina Ivanova³⁷, Stephen B. Milne³⁸

Membership and Editorial Committee: Patrick J. Casey³⁰ (Chair), T. Kendall Harden³⁸

Bridging Project Investigators: Adam P. Arkin¹⁵, John Doyle⁸, Martha L. Gray³⁹, Tobias Meyer¹⁹, Stephen Michnick⁴¹, Martin A. Schmidt⁴⁰, Mehmet Toner⁴², Roger Y. Tsien⁴³

Data Analysis: Madhusudan Natarajan¹, Rama Ranganathan¹, Gilberto R. Sambrano¹²

Departments of ¹Pharmacology, ²Physiology, ³Pathology, ⁴Biochemistry, and ⁵Molecular Biology and Oncology, University of Texas Southwestern Medical Center, Dallas, Texas 75390, USA; ⁶Biology Division, 147-75, ⁷Department of Biology, Howard Hughes Medical Institute, and ⁸Electrical Engineering, California Institute of Technology, Pasadena, California 91125, USA; Departments of ⁹Cellular & Molecular Pharmacology, ¹⁰Microbiology and Immunology, ¹¹Medicine, and ¹²Cardiovascular Research Institute, University of California at San Francisco, San Francisco, California 94143, USA; ¹³Aventis Pharmaceuticals, Bridgewater, New Jersey 08807, USA; ¹⁴Pharmacology, Physiology & Biological Chemistry Division, NIGMS, NIH, Bethesda, Maryland 20892, USA; ¹⁵Departments of Bioengineering & Chemistry, University of California, E.O. Lawrence Berkeley Natl. Laboratory, Berkeley, California 94720, USA; ¹⁶Department of Cell Biol. & Anatomy, Johns Hopkins University, Baltimore, Maryland 21205, USA; ¹⁷Center for Clinical Science Research, Departments of ¹⁸Genetics and ¹⁹Molecular Pharmacology, Stanford University School of Medicine, California 94305, USA; ²⁰Department of Molecular Biology & Biochemistry, University of California, Irvine, California 92697, USA; ²¹Department of Microbiology and Immunology, University of British Columbia, Vancouver, British Columbia, V6T 1Z3 Canada; ²²The Babraham Institute, Cambridge CB2 4AT, UK; ²³Departments of Cell Biology & Medicine, Harvard Medical School, Beth Israel Deaconess Medical Center, Boston, Massachusetts 02215, USA; ²⁴Department of Pharmacology, University of Washington School of Medicine, Seattle, Washington 98195, USA; ²⁵Boston University, BMERC, Boston, Massachusetts 02215, USA; ²⁶Department of Cell Biology, Harvard Medical School, Boston, Massachusetts 02115, USA; ²⁷Department of Biol. Chem., University of Michigan Medical School, Ann Arbor, Michigan 48109, USA; ²⁸The Salk Institute for Biological Studies, San Diego, California 92186, USA; ²⁹Howard Hughes Medical Institute, and ³⁰Department of Pharmacology & Cancer Biology, Duke University Medical Center, Durham, North Carolina 27710, USA; ³¹Mount Sinai Hospital, Toronto, Ontario M5G 1X5, Canada; ³²Memorial Sloan Kettering Cancer Center, New York 10021, USA; ³³San Diego Supercomputer Center, University California, San Diego, La Jolla, California 92037, USA; Department of ³⁴Rheumatology and ³⁵Cardiology, Veterans Affairs Medical Center, San Francisco, California 94121, USA; ³⁶Department of Pharmacology, Vanderbilt University, Nashville, Tennessee 37232, USA; ³⁷118 Hatherley Road, Syracuse, New York 13224, USA; ³⁸Department of Pharmacology, School of Medicine, The University of North Carolina, Chapel Hill, North Carolina 27599, USA; ³⁹Department of Electrical and Medical Engineering, Harvard-MIT Division of Health Sciences & Technology, and ⁴⁰Department of Electrical Engineering & Computer Science, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139, USA; ⁴¹Department of Biochemistry, University of Montreal, Quebec, Canada H3C 3J7; ⁴²Shriners Burn Hospital, Boston, Massachusetts 02114, USA; ⁴³Howard Hughes Medical Institute, University of California at San Diego, La Jolla, California 92093, USA

Unravelling the signal-transduction network in B lymphocytes

Gilberto R. Sambrano*, Grisha Chandyt, Sangdun Choi‡, Dianne Decamp§, Robert Hsueh§, Keng-Mean Lin§, Dennis Mock¶, Nancy O'Rourke†, Tamara Roach||, Hongjun Shu§, Bob Sinkovits¶, Mary Verghese† & Henry Bourne*

*Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143, USA

†Department of Molecular Pharmacology, Stanford University School of Medicine, 975 California Avenue, Palo Alto, California 94305, USA

‡Biology Division, 147-75, California Institute of Technology, Pasadena, California 91125, USA

§Department of Pharmacology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA

¶San Diego Supercomputer Center, University of California, San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

||Department of Rheumatology, San Francisco Veterans Administration Medical Center, 4150 Clement Street, San Francisco, California 94121, USA

The Alliance for Cellular Signaling has chosen the mouse B lymphocyte as a model system to understand basic principles that govern cellular signalling. Progress to that end has focused initially on establishing a reproducible experimental cell system and characterizing essential signalling responses. Although unravelling this complex network will take years, findings revealed in the interim will prove immensely useful to the scientific community at large.

The B lymphocyte has an essential role in host defence and maintains a large complement of cell-surface receptors that regulate its many immunological functions. B lymphocytes travel through the circulation and migrate constantly in and out of the spleen and lymph nodes, ready to act when they encounter antigen, cytokines, bacterial endotoxin, T-cell surface molecules or other stimuli^{1–3}. Migration is directed by a cooperative set of chemokines, and survival is promoted by other environmental stimuli^{4–6}. They may proliferate and differentiate into antibody-secreting plasma cells when stimulated by antigen, and undergo programmed cell death in the absence of stimulation⁷. Scientists have already begun to unlock the key steps in signalling pathways that direct these important functions.

The Alliance for Cellular Signaling (AfCS) chose to study resting B lymphocytes because they possess interesting biological properties regulated by interacting signalling systems. But our goal is not to recapitulate findings that have stimulated immunologists for decades. Rather, we want to collect data from complex signalling pathways in a carefully controlled and defined cell system to address questions about the basic principles underlying signalling mechanisms.

Characterizing the cells

Achieving reliable, experimentally derived measurements from these cells requires the development of a standardized procedure for isolating and purifying splenic B lymphocytes. Borrowing from established methods⁸, investigators at AfCS laboratories devised and refined a reproducible procedure for isolating splenic cells and studying their responses to ligands *in vitro*. As detailed in the first AfCS research report (see www.signaling-gateway.org/reports/v1/BC0001/BC0001.htm), the isolation procedure produced a cell population that is >95% B lymphocytes, of which >80% are resting, mature B2 cells. Cells isolated in this manner maintain their viability over short-term culture and demonstrate reproducible functional responses *in vitro* to an array of 'sentinel ligands', which include anti-immunoglobulin M (anti-IgM) antibody, terbutaline, anti-CD40 antibody, stromal cell-derived factor-1 α (SDF-1 α) and interleukin (IL)-4.

The short lifespan of splenic B lymphocytes in culture limits application of many tools essential for studying signalling networks, including expression of foreign DNA and retroviral strategies for expressing RNA-interference constructs. For this reason, we used the

WEHI-231 cell line as an experimental surrogate. Although no cell line is a perfect substitute, WEHI-231 cells respond to many ligands that act on primary B lymphocytes and are readily amenable to transient and stable transfection of recombinant constructs^{9–12}. So far we have demonstrated robust signalling responses in WEHI-231 cells that are similar to those observed for splenic B lymphocytes. WEHI-231 cells also migrate in response to chemokines and undergo apoptosis in response to B-cell receptor (BCR) stimulation.

We expect WEHI-231 cells to serve as useful adjuncts to our studies of signalling networks of B lymphocytes and as a primary experimental system for the 'Focus on PIP3 modules' (FPM) project, which is designed to accelerate analysis of signalling networks by investigating the relatively small subset of the network that regulates concentrations of the membrane lipid, phosphatidylinositol 3,4,5-trisphosphate (PtdIns(3,4,5)P₃ or PIP₃). However, these cells will impose certain limits on our ability to study desired signalling pathways. For example, the functional response to BCR stimulation in WEHI-231 (apoptosis) differs from that of splenic B lymphocytes (proliferation)^{13,14}. Although the basic organization of most of the WEHI-231 signalling network will probably resemble that of splenic B lymphocytes, analysis of specific pathways may require us to turn to other cells. Accordingly, we are exploring other possibilities, including using splenic B lymphocytes whose life in culture is extended by stimulation with agents that promote survival, such as BAFF or CD40 ligand^{5,6}. Alternatively, we can isolate cells from mice that specifically express the human Bcl-2 transgene in their B-lymphocyte population to promote survival¹⁵. If such cells prove amenable to transfection, they will allow us to study portions of the B-lymphocyte network in addition to those that are retained from the resting state.

Estimating the complexity of the system

Having developed a relatively well-defined system for study, we are now beginning to examine the complexity of the B-lymphocyte signalling network, including identifying the components that participate in the network, identifying interactions that might occur among them, and characterizing the breadth of signalling responses.

Gene-expression profiles

To gain initial insight into the spectrum of signalling components that are expressed in our population of resting B lymphocytes, we have established a collaborative effort with the Genomics Institute of the Novartis Research Foundation¹⁶. A profile of gene expression has been

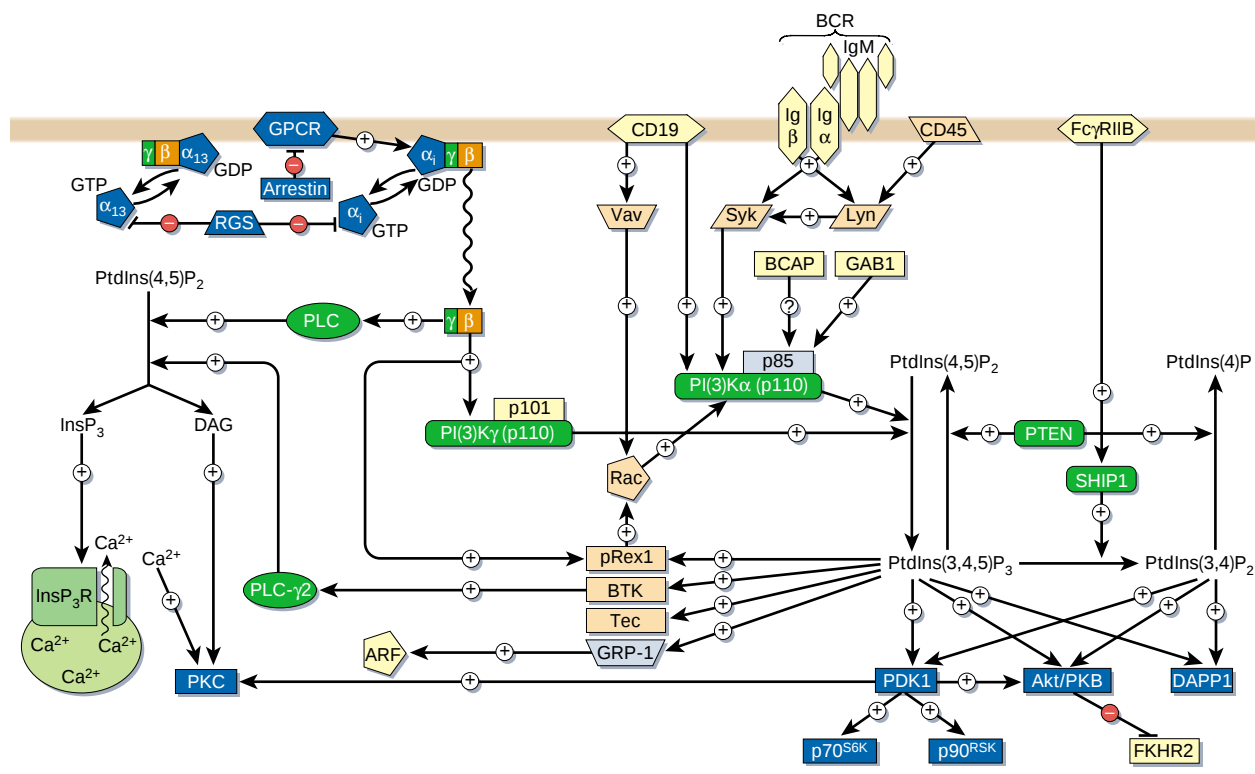


Figure 1 Graphical representation of the PIP₃ signalling module in B lymphocytes.

made using mouse Affymetrix GeneChip arrays which has estimated expression patterns of some 12,000 transcripts. The entire profile can be viewed online at www.signaling-gateway.org/reports/v1/BC0001/SupplGNFArray.xls. These data will supplement additional large-scale transcriptional profiles being compiled within the AfCS laboratories. More detailed analysis of this collection will allow us to focus attention on signalling proteins that are likely to participate in the signalling network of the B lymphocyte.

Protein–protein interactions

In collaboration with Myriad Genetics, we are beginning to compile information on protein–protein interactions from a high-throughput yeast two-hybrid screening system. So far, ~160 signalling proteins have been selected and over 600 bait–DNA-binding-domain hybrids have been prepared for screening. Currently, ~317 interactions from 33 baits have been identified and listed on our web site. This small sample already includes several well-established interactions, including those of phosphatidylinositol-3-OH kinase (PI(3)K) p110α catalytic subunit with the p85α regulatory subunit¹⁷, and the serine/threonine kinase Pak with the small G protein Rac¹⁸. Several new interactions were identified, including those of phospholipase Cγ with the lipid phosphatase SHIP1, and of the tyrosine kinase Btk with the protein phosphatase calcineurin. This ongoing project promises to provide us with a wealth of new information on protein interactions.

Identifying signalling pathways

The alliance's most extensive experimental effort so far has been identifying some of the multiple signalling pathways stimulated in splenic B lymphocytes in response to each of 32 soluble stimulating ligands. The ligands, identified from surveying the literature, include chemokines, a BCR stimulus, lipids, biogenic amines, and multiple interleukins and cytokines that are known or suspected to have effects on B lymphocytes. The goal of this initial screen was to determine which ligands give functionally unique responses across a panel of signalling assays. Experimental responses measured include:

changes in intracellular concentrations of calcium and cyclic AMP, immunoblots of antiphospho antibodies against ten distinct signalling proteins, and expression of approximately 10,000 mouse genes using an Agilent complementary DNA microarray. These assays were chosen to sample a broad range of signalling pathways present in these cells; to these we will add analysis of ligand-induced changes in concentrations of glycerophospholipids and we will expand the panel of antiphospho antibodies.

So far, 18 of 32 ligands showed a positive measurable response in one or more assays, which is consistent with published studies describing the signalling behaviour of these ligands. As expected, ligands known to act primarily on G_s-coupled receptors induced increases in cAMP (with some kinetic differences), but showed no effect on calcium concentrations. Chemokines, which stimulate G_i, all produced kinetically similar increases in intracellular calcium that were accompanied by proportional increases in phosphorylation of Erk1/Erk2 and other targets. A transient increase in cAMP, previously reported for some chemotactic factors¹⁹, was also observed with most of the chemokines that act on B lymphocytes. Changes in gene expression, some substantial, have been detected with virtually all ligands, but more experimentation and analysis will be necessary to distinguish small changes from background noise.

Anti-IgM, CD40 ligand and IL-4 produced many changes in gene expression that are concordant with their varied biological effects on resting B lymphocytes, including proliferation, survival, differentiation and cytokine release. Chemokines produced relatively fewer changes, which is also consistent with their more focused role in directing cell migration. Among the 14 ligands that did not elicit detectable responses in our panel of assays is BAFF, which shows clear proliferative and pro-survival effects on cultured B lymphocytes. BAFF acts on a group of receptors related to tumour-necrosis factor (TNF) receptor, which stimulate the TRAF family of signalling components²⁰. Signalling through this class of receptors typically results in activation of nuclear factor-κB and/or Jnk^{3,20}. Thus our panel of assays is failing to detect signalling pathways triggered by BAFF (and

perhaps by other related ligands, such as TNF- α). The importance of these ligands in B-lymphocyte biology warrants the addition of screening assays and antiphospho antibodies that may permit detection of such signalling responses.

Despite the small number of assays and phosphoproteins examined so far, the overall data set is information-rich, and more extensive analysis will enhance its value. We are just beginning the next experimental step in assessing signal complexity, that of measuring responses to combinations of ligands. The goal will be to detect responses to pairs of ligands that are not geometrically additive sums of responses to the individual ligands. We hope to discover new synergies or inhibitory interactions between ligands, in addition to confirming those already documented. The level of complexity is likely to be great and it will take time to amass the data necessary to begin modelling an entire cellular signalling system. Within the next two years, however, focusing our resources on a more limited and well-defined portion of the signalling network offers a more practical approach to the challenge of network analysis.

A small-scale, detailed analysis of PIP₃ signalling networks

We decided to focus on PIP₃ to recapitulate on a smaller scale the experimental aims of the AfCS (see introductory article on pages 703–706), in part because PIP₃ offers opportunities to monitor signalling processes not only temporally, but also spatially (at the level of individual cells and subcellular compartments)²². This approach is more informative than the measurements made with cell populations, which offer only averages of many individual events.

To bring the network to an experimentally manageable size, we choose to assess only short-term responses (<20 min) to two stimuli (a chemokine, SDF-1 α or BLC, and a BCR stimulus) and to analyse only the signalling network upstream of PIP₃ accumulation (except that phosphorylation of PIP₃-regulated downstream proteins such as Akt/protein kinase B will serve as quantitative readouts for PIP₃ accumulation). Although these stringent limitations will prevent the FPM from considering potentially critical feedback loops downstream of PIP₃, they do define a target network of 100–200 proteins (Fig. 1). This is a small enough network to be manageable and large enough to make it clear that the FPM is not competing with individual laboratories. The proposed FPM project will be carried out in WEHI-231 cells.

Identifying a focal point within the vast cellular network of B lymphocytes provides a basis for prioritizing and directing development of specific 'Molecule Pages', acquiring reagents and extending experiments already under way (for example, yeast two-hybrid screens). The alliance will progress towards its larger objectives while gathering data necessary to implement the FPM project. Our literature-based parts list, for example, has been focused to highlight key players in the chemokine and BCR signalling cascades. The list presently includes proteins that synthesize or degrade phospholipids, as well as receptors, adapters, scaffolds, G-protein components, phosphoproteins and other signalling molecules.

On the experimental front, the FPM project will aim to develop assays that directly measure PIP₃ activity. Additionally, because models of network behaviour depend critically on measuring flow of information through network nodes in time and space, one of the main concerns of the FPM will be to develop assays for activities of intermediate proteins within the PIP₃ signalling network.

These strategies will include the following four analyses. First, phosphorylation of key network proteins will be assessed by immunoblotting with phospho-specific antibodies. In collaboration with Cell Signaling Technology, AfCS researchers have begun testing dozens of such antibodies, directed (for example) against enzymes that synthesize or degrade PIP₃ (including the p85 regulatory component of the PI(3)Ks, PTEN and SHIP1). Second, protein–protein associations will be analysed in response to ligand stimulation. AfCS investigators are constructing retrovirally encoded pulldown baits for proteins in the PIP₃ network. Associated proteins will be detected

by two-dimensional gel analysis and mass spectroscopy, and by immunoblotting. This class of intermediate responses will also help expand the parts list and modify the virtual epistasis map. Third, using green-fluorescent protein tags, the subcellular distribution of all FPM proteins will be assessed. Robust ligand-dependent translocation of a protein between compartments will serve as a useful intermediate endpoint. Finally, fluorescence resonance energy transfer (FRET) probes, which can measure associations of proteins in time and space, will target FPM proteins whose activities are especially critical for information flow through the network.

Whither from here?

Like most good experiments, our exploration of the B-lymphocyte signalling network requires considerable preparation and will take longer than we had hoped. It is too early to predict the experiment's outcome, but the first steps augur well. The project has: devised a convenient experimental cell system that produces reproducible results; begun to assess complexity of outputs to multiple ligands, alone and in combination; begun to develop a comprehensive 'parts list' for signalling in B lymphocytes; and laid groundwork for constructing epistasis maps and assessing intermediate endpoints. Finally, we have embarked on a focused but nonetheless ambitious effort to delineate and dissect a small portion of the signalling network of B lymphocytes. Scientists will enjoy continuing access to every aspect of the experiment on the web and will, we hope, profit from its results. □

doi:10.1038/nature01305

- Hasler, P. & Zouali, M. B cell receptor signaling and autoimmunity. *FASEB J.* **15**, 2085–2098 (2001).
- DeFranco, A. L. The complexity of signaling pathways activated by the BCR. *Curr. Opin. Immunol.* **9**, 296–308 (1997).
- Bishop, G. A. & Hostager, B. S. B lymphocyte activation by contact-mediated interactions with T lymphocytes. *Curr. Opin. Immunol.* **13**, 278–285 (2001).
- Cyster, J. G. Chemokines and cell migration in secondary lymphoid organs. *Science* **286**, 2098–2102 (1999).
- Rollink, A. G., Tschopp, J., Schneider, P. & Melchers, F. BAFF is a survival and maturation factor for mouse B cells. *Eur. J. Immunol.* **32**, 2004–2010 (2002).
- Klaus, G. G., Choi, M. S., Lam, E. W., Johnson-Leger, C. & Cliff, J. CD40: a pivotal receptor in the determination of life/death decisions in B lymphocytes. *Int. Rev. Immunol.* **15**, 5–31 (1997).
- Tarakhovskiy, A. Antigen receptor signalling in B cells. *Res. Immunol.* **148**, 457–460 (1997).
- Miltenyi, S., Muller, W., Weichel, W. & Radbruch, A. High gradient magnetic cell separation with MACS. *Cytometry* **11**, 231 (1990).
- McLeod, S. J., Li, A. H. Y., Lee, R. L., Burgess, A. E. & Gold, M. R. The Rap GTPases regulate B cell migration towards the chemokine SDF-1 (CXCL12). Potential role for Rap2 in promoting B cell migration. *J. Immunol.* **169**, 1365–1371 (2002).
- Reif, K. & Cyster, J. G. RGS molecule expression in murine B lymphocytes and ability to down-regulate chemotaxis to lymphoid chemokines. *J. Immunol.* **164**, 4720–4729 (2000).
- Jakway, J. P., Usinger, W. R., Gold, M. R., Mishell, R. I. & DeFranco, A. L. Growth regulation of the B lymphoma cell line WEHI-231 by anti-immunoglobulin, lipopolysaccharide, and other bacterial products. *J. Immunol.* **137**, 2225–2231 (1986).
- Krebs, D. L., Yang, Y., Dang, M., Haussmann, J. & Gold, M. R. Rapid and efficient retrovirus-mediated gene transfer into B cell lines. *Methods Cell Sci.* **21**, 57–68 (1999).
- Scott, D. W., Livnat, D., Pennell, C. A. & Keng, P. Lymphoma models for B cell activation and tolerance. III. Cell cycle dependence for negative signaling of WEHI-231 B lymphoma cells by anti- μ . *J. Exp. Med.* **164**, 156–164 (1986).
- Benhamou, L. E., Cazenave, P.-A. & Sarthou, P. Anti-immunoglobulins induce death by apoptosis in WEHI-231 B lymphoma cells. *Eur. J. Immunol.* **20**, 1405–1407 (1990).
- Strasser, A. et al. Enforced BCL2 expression in B-lymphoid cells prolongs antibody responses and elicits autoimmune disease. *Proc. Natl Acad. Sci. USA* **88**, 8661–8665 (1991).
- Su, A. I. et al. Large-scale analysis of the human and mouse transcriptomes. *Proc. Natl. Acad. Sci. USA* **99**, 4465–4470 (2002).
- Klippel, A., Escobedo, J. A., Hu, Q. & Williams, L. T. A region of the 85-kilodalton (kDa) subunit of phosphatidylinositol 3-kinase binds the 110-kDa catalytic subunit in vivo. *Mol. Cell Biol.* **13**, 5560–5566 (1993).
- Nisimoto, Y. & Ogawa, H. Interaction between p21-activated protein kinase and Rac during differentiation of HL-60 human promyelocytic leukemia cell induced by all-trans retinoic acid. *Eur. J. Biochem.* **269**, 2622–2629 (2002).
- Iannone, M. A., Wolberg, G., & Zimmerman, T. P. Chemotactic peptide induces cAMP elevation in human neutrophils by amplification of the adenylate cyclase response to endogenously produced adenosine. *J. Biol. Chem.* **264**, 20177–20180 (1989).
- Marsters, S. A. et al. Interaction of the TNF homologues BlyS and APRIL with the TNF receptor homologues BCMA and TACI. *Curr. Biol.* **10**, 785–788 (2000).
- Serunian, L. A., Auger, K. R. & Cantley, L. C. Identification and quantification of polyphosphoinositides produced in response to platelet-derived growth factor stimulation. *Methods Enzymol.* **198**, 78–87 (1991).
- Hurley, J. H. & Meyer, T. Subcellular targeting by membrane lipids. *Curr. Opin. Cell Biol.* **13**, 146–152 (2001).

Correspondence and requests for materials should be addressed to G.R.S. (e-mail: sambrano@cmp.ucsf.edu).

Navigating the signalling network in mouse cardiac myocytes

Gilberto R. Sambrano*, Iain Fraser†, Heping Han‡, Yan Ni‡, Tim O'Connell§, Zhen Yan‡|| & James T. Stull¶

*Department of Cellular and Molecular Pharmacology, University of California, San Francisco, 513 Parnassus Avenue, San Francisco, California 94143, USA

†Biology Division, 147-75, California Institute of Technology, Pasadena, California 91125, USA

Departments of ‡Pharmacology and ¶Physiology, University of Texas Southwestern Medical Center, 5323 Harry Hines Boulevard, Dallas, Texas 75390, USA

§Department of Cardiology, San Francisco Veterans Administration Medical Center, 4150 Clement Street, San Francisco, California 94121, USA

||Present address: Duke University, Durham, North Carolina 27710, USA

Cardiac myocytes have a complex network of signals that regulates their essential role in the rhythmic pumping of the heart. This network is an appealing model system in which to study the basic principles underlying cellular signalling mechanisms. Progress in this effort has come through the establishment of standardized myocyte isolation and culture procedures and characterization of important signalling responses.

The heart is an adaptive organ for pumping blood, responding to changing needs by modifying contractile strength and beating rate¹. The cardiac myocyte is the principal cell in the heart; it coordinates contraction and has the capability to sense a large number of hormonal, neural, electrical and mechanical inputs through a variety of cell-surface and nuclear receptors^{2–9}. Myocytes are also targets of an extraordinary number of physiological and pharmacological agents, because of the critical need to regulate contraction strength and heart rate, and their importance in several cardiovascular diseases.

Signals transmitted by G-protein-dependent and other pathways form a highly cooperative network of interacting molecules¹. This signalling network regulates many complex functions that maintain the rhythmic pumping of the heart, but can also give rise to pathological states, particularly myocardial hypertrophy and failure.

The cardiac myocyte increases in size (hypertrophy) as a compensatory adaptive response, much as skeletal muscle grows with progressive weight training^{10,11}. Cardiovascular diseases, such as hypertension or valvular disorders, increase ventricular wall stresses and trigger a hypertrophic response that is initially compensatory, but the myocardium eventually decompensates (causing heart failure), a common and extremely costly clinical problem¹². Mechanisms responsible for both physiological and pathophysiological hypertrophy are influenced by Ca²⁺-dependent and additional signalling pathways. Indeed, recent observations have linked heart failure to relatively common polymorphisms in genes encoding β_1 - and α_{2c} -adrenergic receptors¹³. Progression from hypertrophy to heart failure can further result in cardiac myocyte apoptosis^{10–12}.

Understanding the relationships between signalling pathways that direct these distinct responses may reveal how a physiological event becomes pathological. It is for this reason that the Alliance for Cellular Signaling (AfCS) has chosen the adult mouse cardiac myocyte as a model system to investigate how cells interpret signals. Knowledge gained using this model will help answer some of the more fundamental questions concerning cell signalling networks (see accompanying introductory article, pages 703–706). Our aim is to understand how varied but linked responses are connected in a complex network of signalling pathways in this cell type.

A cardiac myocyte model system

Although the cardiac myocyte provides an attractive cell model for study, it brings unique challenges. As the primary contractile cell in the heart it is highly specialized and responds to a remarkably diverse spectrum of stimuli. Stimulated by the early observations that cyclic

AMP mediates β -adrenergic stimulation of contractility¹⁴, many physiological and biochemical studies have explored signalling mechanisms in the heart. Myocytes isolated from several species have been the subject of extensive signalling studies, but use of a mouse model is only a recent advance.

New tools have permitted the phenotypic characterization of intact mice and isolated hearts after ablation of genes, or mutation or overexpression of critical proteins. These approaches have provided enormous insights into mechanisms of cardiac development, excitation–contraction (EC) coupling (cell contraction upon electrical stimulation), receptor-mediated signal transduction, and myocardial hypertrophy. But specific conclusions about the underlying signalling mechanisms that contribute to these organismal processes lack certainty owing to modifying haemodynamic and hormonal influences in addition to important interactions between myocytes and interstitial cells. An isolated and defined model system of adult mouse cardiac myocytes is required to address more specific questions about cellular mechanisms without attempting to mimic all of the complexities of the *in vivo* situation.

To adopt the myocyte model for the study of signalling networks, we have defined an experimental plan, established a reproducible

Box 1

Functional responses of mouse myocytes in culture

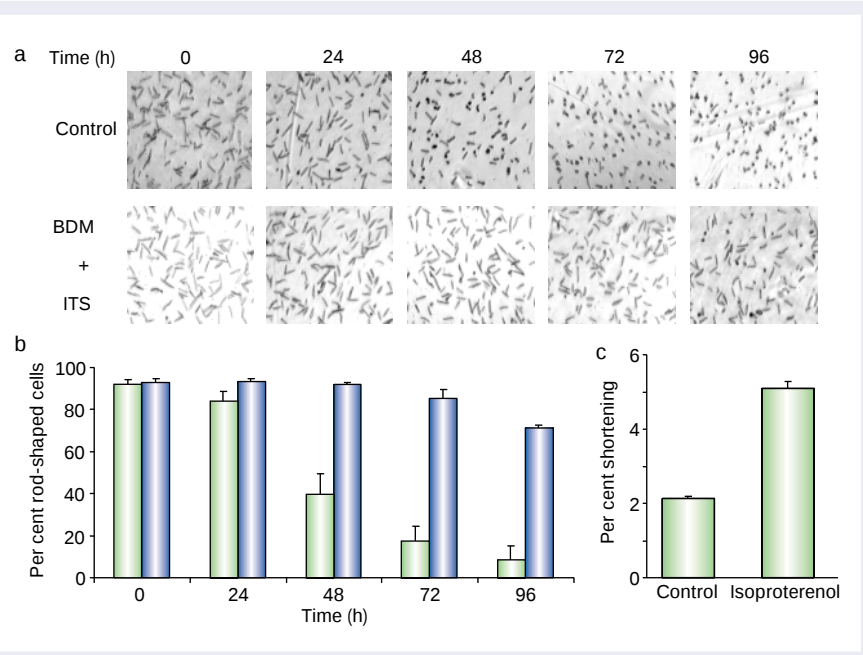
Mouse myocytes cultured for 24 h maintain normal signalling responses, including:

- Contraction with electrical stimulation, thus demonstrating Ca²⁺-dependent excitation–contraction coupling.
- Isoproterenol-induced cell shortening (isoproterenol is a β -adrenergic agonist).
- Isoproterenol-induced cyclic AMP accumulation and phosphorylation of phospholamban (a regulatory protein of the Ca²⁺-ATPase pump in the sarcoplasmic reticulum)^{8,20}.
- Inhibition of isoproterenol-induced cyclic AMP accumulation by carbachol (a muscarinic cholinergic agonist).
- Insulin- and (to a lesser extent) insulin-like growth factor-I-induced phosphorylation of Akt at Ser 473 (refs 21,22).



AFCS CELL PREPARATION LABORATORY

Figure 1 Long-term culture of cardiac myocytes. Adult mouse ventricular myocytes were in serum-free medium (modified minimal essential medium with Hanks' balanced salt solution) for the indicated times without or with 10 mM 2,3 butanedione monoxime (BDM) and 1 $\mu\text{g ml}^{-1}$ insulin, 0.55 $\mu\text{g ml}^{-1}$ transferrin and 0.5 ng ml^{-1} selenium (ITS). **a**, Representative microscopic images of myocytes. **b**, Averaged per cent rod-shaped cells in culture without (green bars) or with (blue bars) BDM and ITS. **c**, At 72 h, six myocytes were stimulated electrically for measurements of per cent cell shortening in the absence and presence of 1 μM isoproterenol for 5 min.



procedure for the isolation and maintenance of mouse cardiac myocytes in culture, and begun characterizing the varied signalling responses in these cells.

Isolation and culture of cardiac myocytes

Primary cells isolated from intact heart have been an important model for study because there are no cell lines that maintain the unique rod-shaped morphology and complement of proteins necessary for EC coupling. In serum-free culture, adult cardiac myocytes from guinea pigs, rats and rabbits are usually quiescent and retain their viability and unique rod-shaped morphology for at least a few days. These cells maintain highly organized membrane and myofibrillar structures that support contractions induced by electrical or pharmacological stimulation, and are amenable to viral-mediated expression of exogenous proteins^{15–17}. But similarly successful culture of mouse cardiac myocytes has been more challenging, perhaps because of difficulties in enzymatic isolation of healthy myocytes and unique variables for relatively long-term culture. As a consequence, less is known about mouse cardiac myocyte physiology.

At the time the AfCS effort was launched, several investigators reported progress in isolation and culture of mouse cardiac myocytes^{18,19}. Given the current focus on mouse genetics, a mouse model offers significant advantages. Our initial goal was therefore to establish procedures for isolation of healthy rod-shaped myocytes that could be maintained in culture for 24 h, thereby providing a population of cells suitable for studies on short-term responses to ligands. In addition, procedures for maintaining myocytes for 72 h or longer were needed to allow manipulation of gene expression in culture using antisense oligonucleotides and RNA interference. These rod-shaped myocytes must retain EC coupling mechanisms and responses to receptor activation, particularly protein phosphorylation events that affect contractility and hypertrophy^{2–9}.

The AfCS has established a standardized procedure for the isolation of ventricular cardiac myocytes from adult mice based on modifications of previously reported protocols^{18,19}. Investigators at the AfCS Laboratory for Development of Signaling Assays focused on initial development of isolation protocols and assessments of pathway responses by protein phosphorylation with phosphospecific antibodies. Meanwhile, researchers at the AfCS Cell Preparation and Analysis Laboratory evaluated EC coupling responses by measuring changes in cytosolic Ca^{2+} and contraction with electrical stimulation. Merging these efforts resulted in a reliable and reproducible method

that yields cells of quantity and quality sufficient for signalling studies in culture. We obtain about one million rod-shaped myocytes per heart, of which 80% are rod-shaped when freshly isolated. A high percentage of cells maintain a functional, rod-shaped morphology after recovery over 24 h, as well as extended culture for up to 72 h (90% and 75% of cells, respectively). Experimental details will be published soon on the Signaling Gateway (www.signaling-gateway.org).

Characterization of cellular responses

Cells cultured in this way exhibit important responses that provide evidence of retained *in vivo* functional attributes as well as suitable signalling endpoints for future study (see Box 1). Additionally, few changes in messenger RNA expression profiles are observed between myocytes cultured for 24 h and freshly isolated cells. These results provide confidence that our model system is ready for additional study and we can begin to assess the complexity of the signalling network. To that end, a broader screening of signalling responses to approximately 30 ligands will begin soon with measurements of short-term changes in cAMP accumulation, cytoplasmic Ca^{2+} concentrations, protein phosphorylation, contraction, and gene expression. These measurements will provide a spectrum of responses for comparison of individual ligands and for detection of interactions between combinations of ligands.

Characterization of functional responses after extended times in culture is also now possible. Investigators in the AfCS laboratories observed that a chemical used routinely in the isolation of cardiac myocytes, 2,3-butanedione monoxime (BDM), helps to maintain rod-shaped, functional myocytes in culture for 72 h or more. BDM is used to inhibit myocyte hyper-contraction during isolation, and during the restoration of Ca^{2+} into the medium. It is also used as a component of cardioplegic solutions to protect from ischaemic damage²³. BDM affects a number of cellular processes including myosin-based contraction, ion currents and Ca^{2+} release²⁴. While BDM or ITS (insulin, transferrin and selenium) alone are not sufficient, the inclusion of both in the culture medium attenuates the rounding of myocytes routinely seen by 48–72 h (Fig. 1).

Although the mechanism(s) for the protective actions are not clear, myocytes cultured for 72 h with both BDM and ITS show normal signalling responses: typical β -adrenergic and muscarinic responses (that is, cAMP accumulation and phospholamban phosphorylation) as well as G_q -coupled phosphorylation and activation

of extracellular signal-regulated kinase. After removal of BDM and ITS for 30 min, most of the cultured myocytes contract in response to electrical stimulation and show a robust positive inotropic response to isoproterenol (a β -adrenergic agonist; Fig. 1). In addition, successful adenovirus-directed expression of a β -galactosidase reporter protein in cultured mouse myocytes indicates that we will be able to manipulate signalling pathways by expressing dominant negative, constitutively active and other interacting and reporter forms of key proteins.

Transcript profiles do change modestly with the extended culture, particularly when comparing freshly isolated cells to those cultured for 72 h; however, the contribution of BDM or ITS to these changes is unclear, because most cells cultured in their absence for 72 h are not viable. Nevertheless, our collective analyses indicate that basic elements of EC coupling and signalling responses seem to be preserved in cardiac myocytes after 72 h in culture with BDM and ITS, making these cells suitable for additional studies after manipulation of gene expression or introduction of exogenous proteins.

Making network connections

To assess the complexity of the cardiac myocyte signalling network, we must compile a 'parts list' of relevant signalling molecules and identify possible interactions among them. Development of a parts list has already begun with a literature-based search of proteins that probably participate in cardiac myocyte signalling and by construction of simple signalling maps focused on insulin, cAMP, Ca^{2+} and phosphatidylinositol 3,4,5-trisphosphate ($\text{PtdIns}(3,4,5)\text{P}_3$ or PIP_3) modules that affect metabolism, gene transcription, hypertrophy and contraction of cardiac myocytes (available at www.afcs.org). The list serves as a source for prioritizing proteins to be used as yeast two-hybrid baits, as tagged proteins for subcellular localization and for quantification, and as targets for perturbations. The list will expand as we find additional proteins that associate with or are regulated by known elements in the network. cDNA and oligonucleotide-based microarrays will be important complements in our effort to construct a comprehensive list and, in time, offer the hope of obtaining detailed information on expression of specific splice variants of individual signalling proteins.

Our current signalling maps provide only a hint of the interactions that may occur within a cellular network. In collaboration with Myriad Genetics, we are beginning to uncover new information on protein-protein interactions using a high-throughput yeast two-hybrid screening system. One example of the results being obtained is an association between histone deacetylase 7 and the transcription factor Mef2c (myocyte enhancer factor-2c), which was previously proposed by Kao and co-workers using immunoprecipitation methods²⁵. Additional novel interactions that may be of special interest include an interaction of the ryanodine receptor with calcium-modulating cyclophilin ligand; the GTPase Dbl with the transcription factor HAND2; and Mef2c with the thyroid hormone receptor interactor 6. This ongoing project will reveal many new protein interactions that can be further analysed by co-immunoprecipitation and fluorescence resonance energy transfer (FRET) measurements.

Another important element in deciphering a signalling network is the ability to measure the flow of information through space and time. A significant portion of this information will be derived from observation of subcellular localization and movements of green fluorescent protein-tagged proteins and examination of FRET probes that permit analysis in real time. The size and complex morphology of cardiac myocytes makes these cells especially interesting subjects for microscopy.

It will be important to integrate dynamic morphological studies of myocytes with information obtained from other AfCS analyses. For example, information obtained from measurements of the translocation of the serine/threonine kinase Akt from cytoplasm to plasma

membrane after activation of transmembrane tyrosine kinase receptors will be integrated with measurements of Akt phosphorylation, as well as with information about other components of systems that generate or respond to PIP_3 . (For a description of the alliance's focus on the PIP_3 module, results from which will also be incorporated into studies on cardiac myocytes, see the accompanying article on pages 708–710.) The insights obtained from these different experimental approaches will both strengthen and expand our appreciation of the network system functioning in myocytes.

Moving ahead

When we launched the AfCS effort two years ago, we could only speculate about our ability to develop a suitable model system using mouse cardiac myocytes. Development of a reproducible culture system for myocytes has been challenging, and although we continue to refine this model, we will soon be characterizing many signalling responses to individual ligands and combinations of ligands. We will identify the connectivity between signalling components, and eventually measure the flow of information through this system as comprehensively as possible. We are confident that the research community at large will profit from the contributions that we will bring to understanding the mouse cardiac myocyte. □

doi:10.1038/nature01306

- Bers, D. M. *Excitation-Contraction Coupling and Cardiac Contractile Force* (Kluwer Academic Publishers, Dordrecht, 2001).
- Koch, W. J., Lefkowitz, R. J. & Rockman, H. A. Functional consequences of altering myocardial adrenergic receptor signaling. *Annu. Rev. Physiol.* **62**, 237–260 (2000).
- Adams, J. W. & Brown, J. H. G-proteins in growth and apoptosis: lessons from the heart. *Oncogene* **20**, 1626–1634 (2001).
- Ross, R. S. & Borg, T. K. Integrins and the myocardium. *Circ. Res.* **88**, 1112–1119 (2001).
- Steinberg, S. F. & Brunton, L. L. Compartmentation of G protein-coupled signaling pathways in cardiac myocytes. *Annu. Rev. Pharmacol. Toxicol.* **41**, 751–773 (2001).
- Xiao, R. P. β -Adrenergic signaling in the heart: dual coupling of the β_1 -adrenergic receptor to G_s and G_i proteins. *Science's STKE* <<http://stke.sciencemag.org/cgi/content/full/sigtrans;2001/104/re15>> (2001).
- Cripps, R. M. & Olson, E. N. Control of cardiac development by an evolutionarily conserved transcriptional network. *Dev. Biol.* **246**, 14–28 (2002).
- Hagemann, D. & Xiao, R. P. Dual site phospholamban phosphorylation and its physiological relevance in the heart. *Trends Cardiovasc. Med.* **12**, 51–56 (2002).
- Rockman, H. A., Koch, W. J. & Lefkowitz, R. J. Seven-transmembrane-spanning receptors and heart function. *Nature* **415**, 206–212 (2002).
- Colan, S. D. Mechanics of left ventricular systolic and diastolic function in physiologic hypertrophy of the athlete's heart. *Cardiol. Clin.* **15**, 355–372 (1997).
- Olson, E. N. & Williams, R. S. Calcineurin signaling and muscle remodeling. *Cell* **101**, 689–692 (2000).
- Hunter, J. J. & Chien, K. R. Signaling pathways for cardiac hypertrophy and failure. *N. Engl. J. Med.* **341**, 1276–1283 (1999).
- Small, K. M., Wagoner, L. E., Levin, A. M., Kardia, S. L. & Liggett, S. B. Synergistic polymorphisms of β_1 - and α_1C -adrenergic receptors and the risk of congestive heart failure. *N. Engl. J. Med.* **347**, 1135–1142 (2002).
- Robison, G. A., Butcher, R. W., Oye, I., Morgan, H. E. & Sutherland, E. W. The effect of epinephrine on adenosine 3',5'-phosphate levels in the isolated perfused rat heart. *Mol. Pharmacol.* **1**, 168–177 (1965).
- Kruppenbacher, J. P., May, T., Eggers, H. J. & Piper, H. M. Cardiomyocytes of adult mice in long-term culture. *Naturwissenschaften* **80**, 132–134 (1993).
- Benndorf, K., Boldt, W. & Nilius, B. Sodium current in single myocardial mouse cells. *Pflügers Arch.* **404**, 190–196 (1985).
- Wolska, B. M. & Solaro, R. J. Method for isolation of adult mouse cardiac myocytes for studies of contraction and microfluorimetry. *Am. J. Physiol.* **271**, H1250–H1255 (1996).
- Hilal-Dandan, R., Kanter, J. R. & Brunton, L. L. Characterization of G-protein signaling in ventricular myocytes from the adult mouse heart: differences from the rat. *J. Mol. Cell. Cardiol.* **32**, 1211–1221 (2000).
- Zhou, Y. Y. *et al.* Culture and adenoviral infection of adult mouse cardiac myocytes: methods for cellular genetic physiology. *Am. J. Physiol. Heart Circ. Physiol.* **279**, H429–H436 (2000).
- Simmerman, H. K. & Jones, L. R. Phospholamban: protein structure, mechanism of action, and role in cardiac function. *Physiol. Rev.* **78**, 921–947 (1998).
- Condorelli, G. *et al.* Akt induces enhanced myocardial contractility and cell size in vivo in transgenic mice. *Proc. Natl Acad. Sci. USA* **99**, 12333–12338 (2002).
- Crackower, M. A. *et al.* Regulation of myocardial contractility and cell size by distinct PI3K–PTEN signaling pathways. *Cell* **110**, 737–749 (2002).
- Thum, T. & Borlak, J. Butanedione monoxime increases the viability and yield of adult cardiomyocytes in primary cultures. *Cardiovasc. Toxicol.* **1**, 61–72 (2001).
- Sellin, L. C. & McArdle, J. J. Multiple effects of 2,3-butanedione monoxime. *Pharmacol. Toxicol.* **74**, 305–313 (1994).
- Kao, H. Y. *et al.* Mechanism for nucleocytoplasmic shuttling of histone deacetylase 7. *J. Biol. Chem.* **276**, 47496–47507 (2001).

Correspondence and requests for materials should be addressed to G.R.S. (e-mail: sambrano@cmp.ucsf.edu).

The Molecule Pages database

Joshua Li*, Yuhong Ning*, Warren Hedley*, Brian Saunders*, Yongsheng Chen*, Nicole Tindill†, Timo Hannay†
& Shankar Subramaniam*‡

*San Diego Supercomputer Center, and ‡Departments of Bioengineering and Chemistry and Biochemistry, University of California at San Diego, 9500 Gilman Drive, La Jolla, California 92093, USA

†Nature Publishing Group, The Macmillan Building, 4 Crinan Street, London N1 9XW, UK

The Alliance for Cellular Signaling (AfCS)–Nature Molecule Pages will be a comprehensive database of key facts about more than 3,000 proteins involved in cell signalling. Each entry will be created by invited experts and be peer-reviewed. Alongside the large-scale experiments being conducted by the AfCS scientists, the wealth of information contained in this database offers the potential of accelerating the pace of discovery in signal transduction research.

The understanding of cellular function and physiology requires insight into the complexity of the proteomic content of cells. The complete sequence of mammalian genomes provides a first sense of this content, but genome annotations do not capture many important aspects that contribute to the vast diversity of protein content and function. Splice variants and small variations in the genomic sequence translate into sequence polymorphisms in the proteome, which can alter protein function and contribute to pathophysiological conditions. More important, proteins display large repertoires of 'states', which contribute to the enormous diversity we observe in biochemical networks.

The state of a signalling molecule is characterized by covalent modifications of the native polypeptide, the substrates or ligands bound to the protein, its state of association with other proteins, and its location in the cell. A signalling molecule may be a receptor, a channel, an enzyme or another functionally defined class, and its state modulates its function. During signal transduction, a molecule may undergo a transition from one functional state to another.

Unlike the 'gene' parts-list of a cell, which can be obtained from high-throughput gene array measurements, the states of a protein can at present be ascertained only by painstaking biochemical and cellular experimentation. Our knowledge of the states of intracellular signalling proteins comes from detailed experiments and comparative analysis of cellular pathways across different species. But how can this vast amount of knowledge present in scientific literature be made easily accessible to the scientific community at large, rather than only to expert researchers whose research focus is the molecule of interest?

The Alliance for Cellular Signaling (AfCS)–Nature 'Molecule Pages' is a comprehensive database that will capture qualitative and quantitative information about a large number of signalling molecules and the interactions between them. It will be freely accessible from the AfCS–Nature Signaling Gateway web site (www.signaling-gateway.org/). The Molecule Pages will contain data from many public repositories in addition to information from published literature entered by expert authors. Authors will construct pages by entering information into web-based forms designed to standardize data input.

One of the principal barriers on constructing a database such as this lies in the complex and varied vocabulary used by biologists to define the attributes of a molecule. The database can be useful only if the information is described with a structured vocabulary along with well-defined relationships between the data 'objects' (for example, a protein sequence and a given modification). The building of this 'schema', or the database structure, is thus the first step towards the structured recording of available data about biochemical networks.

Database design considerations

The best way to structure the immense amount of knowledge associated with the states of a protein is to capture it in a relational database (one

that records data in a series of tables in a way that avoids redundancy). This contains precisely defined data fields and precisely defined relationships between them represented by links between the tables.

The Molecule Pages database contains over 200 such relational tables. These define a myriad of parameters ranging from sequence to kinetic and thermodynamic parameters associated with molecular states and state transformations. The complete schema of the database (also known as the entity relationship diagram) is available on request from the AfCS bioinformatics group.

The Molecule Pages are freely accessible using any common web browser. The underlying system has a typical three-tier architecture. Behind the browser (or 'client') tier is a second ('application') tier composed of a collection of Java and Perl programs developed by the AfCS bioinformatics team and hosted on our servers; these provide the functionality of the system. The third ('data') tier, also hosted by the AfCS, uses an Oracle 9i relational database management system to store the raw data.

Description of the Molecule Pages database

The Molecule Pages database schema is divided into automated and author-entered data. The central part of the database is the molecule table, which defines each of over 3,000 AfCS proteins and forms an 'anchor' for both automated and author-entered data about each molecule. The primary element in the molecule table is the canonical mouse protein sequence, which is defined in an unambiguous manner from the corresponding GenBank sequences.

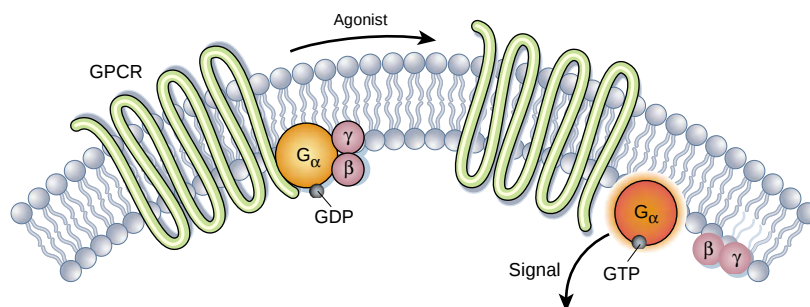
Automated data

The automated data component of each Molecule Page integrates information about that protein obtained from external database records. These include DNA and protein sequence information, structural information, sequence comparisons and related sequences, and basic biophysical and biochemical properties. These data come from SwissProt, GenBank, LocusLink, MGDB (Mouse Genome Database from Jackson Laboratories), Pfam, PIR, PRINTS, TrEMBL, TrEMBLnew, RefSeq, and the Interpro databases. These automated data are stored in relational tables, using the protein GI number (the GenInfo Identifier from GenBank) as the primary key. Links are provided to relevant National Center for Biotechnology Information (NCBI) database records. Each sequence can be imported into the Biology Workbench (<http://workbench.sdsc.edu/>) for further analysis. (The Biology Workbench is a web-based infrastructure that allows biologists to search and analyse many popular protein and nucleic-acid sequence databases.) The automated data are updated on a periodic basis.

Author-entered data

One of the main objectives of the Molecule Pages is to provide the community with information about the function of every protein

Figure 1 Example of molecule state changes. In unstimulated cells, the state of G_α is defined by its interaction with GDP, $G_{\beta\gamma}$, and a G-protein-coupled receptor (GPCR). Upon receptor stimulation by an agonist, its state is changed owing to its dissociation from the receptor and $G_{\beta\gamma}$, and the exchange of GTP for GDP, leading to G_α activation.



associated with cellular signalling networks, including qualitative and quantitative properties of each protein state. Such comprehensive, structured information about the states of signalling molecule proteins will aid the reconstruction — and hence the deeper understanding — of the signalling networks in which they participate. In addition, this information will help to provide the necessary data and parameters for quantitative modelling of signalling networks.

The primary role of an expert author is the entry and curation of these data in the Molecule Pages database. The author-entered data on a given state of the molecule are separated conceptually into two parts: the state of the molecule and the characterization of the function of that state. When describing the state, the author may invoke any number of possible modifications to the native polypeptide. These modifications include: the interaction of the protein with another protein, covalent modifications, binding to a substrate or ligand, and localization in a given subcellular compartment. In most cases a functional state of a protein will involve combinations of the above modifications.

The signalling protein G_α , for example, is in a defined functional state when bound to $G_{\beta\gamma}$, a G-protein-coupled receptor and GDP (Fig. 1). Agonist binding activates the receptor, which results in exchange of the substrate GDP by GTP, and leads to activation of the G protein. The author can define the bound state of G_α by defining the association with nucleotide GDP, $G_{\beta\gamma}$, and the receptor. The molecules with which G_α interacts are in turn defined by various functional states. For example, G_β is constitutively bound to G_γ . Even if no detailed information on $G_{\beta\gamma}$ has been entered previously into the database, the author of the G_α entry will be able to do so in a provisional form. At some time in the future, authors of G_β and G_γ entries will then be able to define the constitutive $G_{\beta\gamma}$ state more fully. In describing the function of G_α in this manner the author can define the activated state where the agonist is bound to the receptor followed by dissociation of G_α from the complexed state.

At each of these steps the author can enter any available quantitative data, such as binding constants or kinetic constants, along with the conditions under which the measurements were made. Detailed formats for entering such quantitative data have been defined in the Molecule Pages database. The above example shows how a complicated functional state of a molecule and its associated properties can be captured in tightly defined and well-structured database format. In this manner, the information can be put to use in the computer-assisted analysis and modelling of cellular signalling networks.

An important part of the Molecule Pages database is the information relating to functional alteration that arises from mutations of the components. For each AfCS molecule, we provide a format for defining all known sequence mutations, including natural variants,

point mutations and deletions or insertions. In the definition of the state of a molecule, the author has the provision to indicate if the mutations alter the functional state and its properties. These data will provide valuable inputs into modelling the changes in cell signalling networks arising from mutations in component molecules and has potential implications in understanding the molecular basis of disease. The Molecule Pages database will also permit the author to provide a citation list pertinent to each entry and will link to the PubMed database maintained by the NCBI at the National Institutes of Health. Every Molecule Page will include an abstract providing a succinct overview of the characteristics and function of the protein in question. After each page has been created, peer reviewed and published, they will be updated annually to ensure that they remain current as more knowledge about a protein is gathered.

Relationship to AfCS experiments

Reconstruction of biochemical pathways is a complex task. In metabolic pathways, the task is somewhat simplified because of the essentially linear nature of the underlying processes, in which each step represents the enzymatic conversion of a substrate into a product. This is not the case in cellular signalling, where effects such as branching and cascades are commonplace.

The role of each protein in a signalling network is to communicate the signal from one node to the next, and to accomplish this, the protein has to be in a particular 'state'. We anticipate that the Molecule Pages will provide a catalogue of states for each significant signalling protein, such that one can begin to reconstruct signalling pathways with molecules in well-defined states functioning as nodes of a network. Interactions within and between functional states of molecules, as well as transitions between functional states, provide the building blocks for reconstruction of a signalling network. As described in the accompanying papers — see introductory article (pages 703–706) and articles on the signalling networks in B lymphocytes and cardiac myocytes (pages 708–710 and 712–714, respectively) — the experiments conducted by AfCS scientists will contribute to testing and validating such interactions and transitions in specific cells of interest. Together with the new large-scale experimental data sets being generated by AfCS laboratories, the highly structured review data provided by the Molecule Pages will, we hope, provide a new foundation for further accelerating the pace of discovery in cellular signalling, thus greatly enhancing our understanding of cellular processes in health and disease. □

doi:10.1038/nature01307

Correspondence and requests for materials should be addressed to S.S. (e-mail: shankar@ucsd.edu).

Contacts

Publisher: Ben Crowe

Editor: Paul Smaglik

Marketing Manager: David Bowen

European Head Office, London

The Macmillan Building
4 Crinan Street
London N1 9XW, UK
Tel +44 (0) 20 7843 4961
Fax +44 (0) 20 7843 4996
e-mail: naturejobs@nature.com

Senior European Sales Manager:
Nevin Bayoumi (4978)

UK/ RoW/ Ireland:

Matt Powell (4953)

Andy Douglas (4975)

Frank Phelan (4944)

Netherlands/ Italy/ Iberia:

Evelina Rubio Hakansson (4973)

Scandinavia: Silje Opstrup (4994)

France/ Belgium:

Amelie Pequignot (4974)

Production Manager:

 Billie Franklin

To send materials use London
address above.

Tel +44 (0) 20 7843 4814

Fax +44 (0) 20 7843 4996

e-mail: naturejobs@nature.com

International

Advertising Coordinator:

Hind Berrada (4935)

Naturejobs web development:

Tom Hancock

Naturejobs online production:

Ben Lund

European Satellite Office

Germany/ Austria/ Switzerland:

Patrick Phelan, Odo Wulffen

Tel + 49 89 54 90 57 11/-2

Fax + 49 89 54 90 57 20

e-mail: p.phelan@nature.com

o.wulffen@nature.com

US Head Office, New York

345 Park Avenue South,
10th Floor, New York, NY 10010-1707

Tel +1 800 989 7718

Fax +1 800 989 7103

e-mail: naturejobs@natureny.com

US Sales Manager:

 Peter Bless

US Advertising Coordinator:

Linda Adam

Japan Head Office, Tokyo

MG Ichigaya Building (5F),

19-1 Harakatomachi,

Shinjuku-ku,

Tokyo 162-0841

Tel +81 3 3267 8751

Fax +81 3 3267 8746

Asia-Pacific Sales Director:

Hideki Watanabe

e-mail: h.watanabe@naturejpn.com

naturejobs

Another British invasion?

The Beatles, The Rolling Stones and The Who took the United States by storm in the 1960s. Pulp, Blur and Suede failed to follow in their footsteps in the 1990s. Now, physicists, chemists, biologists, engineers and mathematicians look to be the next wave of prospective invaders — but as yet, the outcome is uncertain.

There are various signs that the British are heading west. Within the past year, for example, the British Embassy has tripled to 12 the number of science and technology officers posted in the United States. And an organization, British Expats in Life Sciences (www.belsgroup.com), was quietly established in the Boston area. In addition, a survey by the Royal Society found that in 1999, 12% of the society's fellows questioned were working in the United States.

The knee-jerk reaction is that Britain is losing its scientists to a brain drain. But the reality is more complicated. According to one UK government report on science and technology, released this summer, the past few years have seen a wave of scientists head for the United States. But at the same time, Britain experienced a net gain of 5,000 scientists from other countries.

And the exodus across the Atlantic may be helping Britain to establish a technology-transfer beachhead. The results of such endeavours have been mixed. For example, a plan to set up a joint institute between the University of Cambridge and the Massachusetts Institute of Technology was met with optimism, but has recently been criticized for a lack of tangible results (see *Nature* 420, 256; 2002). A less ambitious programme, announced this summer, to foster links between Rice University and Imperial College in London still holds promise. The way to judge such efforts may be not by how many UK programmes and scientists establish links in the United States, but by what they eventually bring back.

Paul Smaglik

Naturejobs editor



Contents

REGIONS

South Korea gets back
into shape

p4

WWW.NATUREJOBS.COM

Career centre
Information on the
scientific job market

FOCUS

SPOTLIGHT

RECRUITMENT

SCIENTIFIC ANNOUNCEMENTS

SCIENTIFIC EVENTS



Fresh horizons South Korea

For the Republic of Korea (South Korea), years of rapid economic growth came at a price. In 1997, the unrestricted lending by Korean banks helped to trigger a cash crisis that was only resolved when the International Monetary Fund agreed to bail the country out to the tune of US\$58 billion. A major factor behind the crisis was the country's dependence on — and overinvestment in — a few key industries, notably the slumping memory-chip market.

Now back on its feet, South Korea is pursuing a policy of tax incentives to spur the growth of high-tech businesses in a wide variety of fields including biotechnology, and is ploughing funds into scientific research in an effort to reinvigorate academia. To aid this transformation, the government, universities and industry are all wooing foreign researchers to gain valuable overseas expertise.

"We are trying to prepare for the future," says Kwan Rim, chairman of the Samsung Advanced Institute of Technology (SAIT) in Suwon. Set up in 1987, the institute is now a cornerstone in Samsung's plans to diversify from its core electronics business. Although 40% of SAIT's 850 researchers are working in digital- and telecommunications-related areas, the organization is confidently building up research programmes in biotechnology and fuel cells.

Petrochemicals, another of South Korea's core industries, is also attempting to diversify, with both LG Chem in Taejeon and Isu Chemical in Seoul seeking fresh markets in biotechnology and drug discovery. But it is a slow process. SAIT's DNA chips, for example, have yet to prove themselves in the Korean market, and LG Chem's new therapeutics are still tied up in clinical trials.

Foreigners, or Koreans who have trained abroad, are seen by many as key to the success of these initiatives. "Pharma is new to us, so we must depend on foreign expertise," says Jong-Kee Yeo, president of LG Chem's Research Park. In an effort to kick-start its biotech

ambitions, LG Chem has opened facilities in the United States, including a site in San Diego that carries out collaborative research with US companies.

This thirst for skilled workers is opening up some reasonably lucrative opportunities, particularly for Korean scientists who have trained abroad. Having gained a taste for the lifestyle and wages offered by the West, many Koreans seem slightly reluctant to return home. "We used to ask: 'Wouldn't it be nice to be in your motherland?'. But that doesn't work any more,"

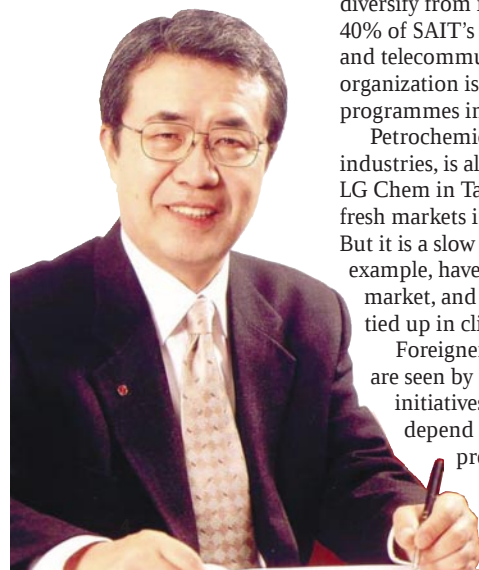
Setting a precedent

Success in biotechnology is not new to South Korea. Macrogen, based in Seoul, made its money from DNA sequencing. Founded as a laboratory venture in 1983 by Jeong-Sun Seo at Seoul National University, Macrogen had plenty of time to develop before it became a company in 1997. Its success hinged on price — it cut the cost of sequencing from US\$15 per read to \$5, which it claims is the lowest in the world. "Our timing was very good," says Seo. The company collared \$40 million when it was listed on Korea's stock exchange and is now

moving into a range of undertakings including a Mongolian genome project, a Korean single nucleotide polymorphism project, creation of knockout mouse models, and DNA-chip production. Younger but also on the rise is Bioneer, a venture founded in 1992. It produces artificial nucleotide sequences often used to make DNA. "These will accelerate genomic studies," says founder and president Han-Oh Park. Using devices of its own design, Bioneer has created its own nucleotide factory, which it claims is much faster than those of its competitors.

D.C.

Korea needs foreign expertise if it is to crack new markets, according to Jong-Kee Yeo.



says Yeo. Instead, LG Chem, Samsung and others are forced to match US salaries, which is especially generous considering South Korea's low taxation rates. To attract foreign researchers, SAIT has built living quarters, fitness facilities and arranged for English-speaking doctors to be available.

Korean companies are also seeking expertise from other sources. Both LG Chem and SAIT plan to more than double, to 50% and 25%, respectively, the amount of research that they outsource to academic researchers in or outside Korea. "We want to gain speed by using the vast reservoirs of talent outside the company," says Rim.

The story is similar at South Korea's academic institutions. The Korea Advanced Institute of Science and Technology (KAIST) in Taejeon, for example, has recruited 27 foreign faculty members to either full-time or visiting positions this year. "They are paid on a different scale, which puts them at or near the salary at their home institution," says KAIST dean of academic affairs Hai-Woong Lee.

This spring, the education ministry also launched a programme to attract foreigners to its universities. In September its first batch of 100 foreign faculty members, including some Koreans who have more than five years' experience abroad, arrived to teach biotechnology and information technology. Sixty-five of the researchers headed for the prestigious Seoul National University. Their salaries, which were negotiated based on their previous position, created a stir as some were as high as US\$100,000 — roughly three times the average salary of a Korean university researcher.

LINGUISTIC LARGESSE

To make foreign researchers feel at home, many institutions are increasingly emphasizing English. Privately run Pohang University of Science and Technology, for instance, says that it wants all of its graduate programmes to be taught in English by 2005.

Such efforts have started to take hold, says Ewan Stewart, a British physicist now at KAIST. Before coming to KAIST, Stewart spent several years in Japan. "Korea seems much more open to foreign researchers, particularly in terms of faculty positions," he says.

But day-to-day living, especially with a family in tow, can be difficult. In Taejeon, a two-hour bus ride south from Seoul and home to LG Chem and KAIST, English is not widely spoken. Communication problems can also make teaching students difficult, says Stewart. "Some undergraduates have great difficulty in lectures and have to rely on my lecture notes on the web."

Many South Korean researchers are nevertheless keen to emphasize the elements of their system that resemble the United States. And there is little doubt that the country's research community has aspects that set it apart from other Asian countries, notably Japan.

One factor is the emphasis on youth rather than a seniority-based hierarchy. "We ensure youth is represented when choosing reviewers," says Chong Shik Chin, director general of the Korea Science and Engineering Foundation in Taejeon. At 61, Chin is a little embarrassed at being the oldest person in the organization. The average age of grant-proposal reviewers is now just under 40, he says.

Industry also has a youthful vigour. "Three years ago we abolished the seniority system," says Rim, adding that the average age of SAIT's researchers is under 40. Efforts are also being made to level the playing field for women (see "Equal opportunities", below).

South Korea is once more a dynamic, fast-moving society and offers a range of opportunities for foreign researchers. If it can maintain its overwhelming energy and build on its successes (see "Setting a precedent", left), it is likely that it will manage to carve out a strong niche in its chosen global markets. ■

David Cyranoski is *Nature's* Asian Pacific correspondent.

Korea Advanced Institute of Science and Technology

♦ www.kaist.edu

Korea Science and Engineering Foundation

♦ www.kosef.re.kr/global

Seoul National University

♦ www.snu.ac.kr/engsnu

Samsung Advanced Institute of Technology

♦ www.sait.samsung.co.kr/sait/src/saitEnIndex.html

LG Chem

♦ www.lgchem.com

Equal opportunities

Last month, the South Korean government passed a law aimed at improving prospects for women scientists. The science and technology ministry must now lay down periodic plans to train and fund women in research. The law also requires the education ministry to produce a plan to recruit women to fields of science more actively.

The Korea Science and Engineering Foundation in Taejeon already has several funding schemes specifically for women, but Chong Shik Chin, director general of the foundation, agrees that in general they are not sufficient to maintain a career. "We all agree that it is a long-term problem that needs to be faced," he says.

There are some positive signs, though. Of the 19 centres set up under South Korea's 21st Century Frontier R&D Program — 10-year programmes in which

researchers form teams focused on specific fields and then turn the results into commercial products — two are directed by women. One is Hyangsook Yoo, who heads the Center for Functional Analysis of the Human Genome. The centre has six females among its 39 team leaders.

"It would be nice to have more women scientists, but we do not choose on the basis of gender," says Yoo. "If these training programmes work, after a few years, maybe we won't need such special attention." D.C.



Hyangsook Yoo is optimistic for the future of women scientists in South Korea.